TRANSCRIPTIONAL REGULATION OF THE HUMAN HYPOXANTHINE PHOSPHORIBOSYLTRANSFERASE GENE BY X CHROMOSOME INACTIVATION

By

IAN KERST HORNSTRA

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1993

I would like to dedicate this dissertation to my parents and family who have graciously supported me through this great endeavor.

ACKNOWLEDGEMENTS

I would like to acknowledge my mentor, Thomas P. Yang, for his enthusiasm and support through the years. I also would like to thank all my friends for their help over the years. In addition, I would like to recognize the members of the Yang lab for many interesting discussions and for technical assistance.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

LIST OF FIGURES vi

ABBREVIATIONS ix

ABSTRACT xi

CHAPTER 1
INTRODUCTION 1
X Chromosome Inactivation 1
Hypoxanthine Phosphoribosyltransferase 9
FMR1 Gene 10
Specific Aims and Rationale 13

CHAPTER 2
MULTIPLE IN VIVO FOOTPRINTS ARE SPECIFIC TO THE ACTIVE ALLELE OF THE X-LINKED HUMAN HYPOXANTHINE PHOSPHORIBOSYLTRANSFERASE GENE 5'REGION: IMPLICATIONS FOR X CHROMOSOME INACTIVATION 16
Introduction 16
Materials and Methods 19
Cell Lines 19
Preparation of DNA--In Vivo Dimethysulfate Treatment and DNA Isolation 21
In Vitro DMS Treatment 23
Ligation-Mediated PCR 24
Gel Electrophoresis and Electrotransfer . . . 26
Probe Synthesis, Hybridization, and Washing . 27
Results 30
Discussion 47
DNA-Protein Interactions Specific to the Active HPRT Allele 48
Comparison of in Vivo Footprinting of Human HPRT and PGK-1 53
Implications for X Chromosome Inactivation . 55

CHAPTER 3
IN VITRO RECONSTITUTION OF A DNA-PROTEIN INTERACTION SPECIFIC TO THE ACTIVE HPRT ALLELE 58
Introduction 58
Materials and Methods 60

Nuclear Extracts 60
Preparation of Cloned DNA Fragments for Gel Mobility-Shift Assays 60
Electrophoretic Gel Mobility Shift Assays . . 63
Results 64
Discussion 70

CHAPTER 4
HIGH RESOLUTION METHYLATION ANALYSIS OF THE HUMAN HYPOXANTHINE PHOSPHORIBOSYLTRANSFERASE GENE 5' REGION ON THE ACTIVE AND INACTIVE X CHROMOSOMES: CORRELATION WITH GENE SILENCING AND BINDING SITES FOR TRANSCRIPTION FACTORS . 74
Introduction 74
Materials and Methods 80
DNA, Cells, and Cell Lines 80
DNA Preparation and Base-Specific Modification 81
Ligation-Mediated PCR 82
Results 85
Analysis of the Lower Strand 89
Analysis of the Upper Strand 103
Summary of Methylation Analysis 105
Discussion 110
Correlation of Cytosine Methylation and the Binding of Transcription Factors . . . 111
Comparison of Cytosine Methylation Patterns on the Human HPRT and PGK-1 Gene 5' Regions 115
Implications for X Chromosome Inactivation 117

CHAPTER 5
HIGH RESOLUTION METHYLATION ANALYSIS OF THE FMR1 GENE TRINUCLEOTIDE REPEAT REGION IN FRAGILE X SYNDROME . . 121
Materials and Methods 125
DNA and Cell Lines 125
DNA Preparation and Base-Specific Modification and Cleavage 125
Ligation-Mediated PCR 126
Results 129
Analysis of the Lower Strand 132
Analysis of the Upper Strand 139
Discussion 142

CHAPTER 6
CONCLUSIONS AND FUTURE DIRECTIONS 148

REFERENCE LIST 156

BIOGRAPHICAL SKETCH 169

LIST OF FIGURES

Figure

2.1 - Location of primers used in the LMPCR analysis of the human HPRT 5' region 33

2.2 - In vivo footprints in the region spanning positions -75 to -98 using primer set E 37

2.3 - In vivo footprints in the region spanning positions -75 to -98 using primer set M 38

2.4 - In vivo footprint analysis of the region spanning positions -159 to -215 using primer set M 41

2.5 - In vivo footprint analysis of the region spanning positions -159 to -215 using primer set C 42

2.6 - In vivo footprint analysis of the region spanning positions -256 to -267 using primer set A 44

2.7 - Summary of in vivo footprint analysis of the human HPRT gene 5' region 46

3.1 - Sequence and restriction Map of human HPRT 5' region used to prepare cloned DNA fragments for gel mobility-shift assays 61

3.2 - Electrophoretic mobility-shift assays using cloned promoter regions fragments from other genes as unlabelled competitor DNA 67

4.1 - Location of primers used in LMPCR genomic sequencing analysis of the Human HPRT 5' region 90

4.2 - Genomic sequencing and methylation analysis of the human HPRT 5' region on the lower strand using primer set N 92

4.3 - Genomic sequencing and methylation analysis of the human HPRT 5' region on the lower strand using primer set A 93

4.4 - Genomic sequencing and methylation analysis of the human HPRT 5' region on the lower strand using primer set M 94

4.5 - Genomic sequencing and methylation analysis of the human HPRT 5' region on the lower strand using primer set I 95

4.6 - Genomic sequencing and methylation analysis of the human HPRT 5' region on the upper strand using primer set J 96

4.7 - Genomic sequencing and methylation analysis of the human HPRT 5' region on the upper strand using primer set E 97

4.8 - Genomic sequencing and methylation analysis of the human HPRT 5' region on the upper strand using primer set C 98

4.9 - Genomic sequencing and methylation analysis of the human HPRT 5' region on the upper strand using primer set R 99

4.10 - Summary of the methylation pattern of cytosines from the human HPRT 5' region on the inactive X chromosome in hybrid cell line 8121 108

4.11 - Summary of the methylation pattern of cytosines from the human HPRT 5' region on the inactive X chromosome in hybrid cell line X8-6T2 109

5.1 - Location of primers used for the genomic sequencing of the human FMR1 gene repeat region 131

5.2 - Genomic sequencing and methylation analysis of the trinucleotide repeat and immediate flanking region on the lower strand using primer set L 134

5.3 - Genomic sequencing and methylation analysis of the trinucleotide repeat and immediate flanking region on the upper strand using primer set U 140

5.4 - Summary of the methylation state of cytosines from the human FMR1 gene repeat region in

affected males and the normal FMR1 gene on the inactive X chromosome 143

ABBREVIATIONS

3'	Three prime
5-azaC	5-azacytidine
5'	Five prime
A	Adenine
ATP	Adenosine Triphosphate
bp	Base pair
C	Cytosine
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytosine triphosphate
dGTP	Deoxyguanosine triphosphate
dNTP	Deoxyribonucleotide triphospate
dTTP	Deoxythymidine triphosphate
DMS	Dimethylsulfate
DNA	Deoxyribonucleic acid
DNase I	Deoxyribonuclease I
EDTA	Ethylenediamine tetra-acetic acid
FBS	Fetal bovine serum
g	one force of gravity
G	Guanine
GMP	Guanosine monophosphate
G6PD	Glucose-6-phosphate dehydrogenase
HAT	Hypoxanthine, aminopterin, thymidine
HPRT	Hypoxanthine phosphoribosyltransferase
IMP	Inosine monophospate
kb	Kilobase pair
LMPCR	Ligation-mediated polymerase chain reaction
ml	Milliliter
mM	Millimolar
M	Molar
mRNA	Messenger ribonucleic acid
PBS	Phosphate-buffer saline
PGK	Phosphoglycerate kinase
PCI	Phenol, chloroform, isoamylalcohol
PCR	Polymerase chain reaction
P-S	Penicillin-streptomycin
pmol	Picomole
RNA	Ribonucleic acid
SDS	Sodium dodecylsulfate
T	Thymidine
TBE	Tris, boric acid, EDTA
TE	10 mM Tris, 1 mM EDTA
Tris	Hydroxymethyl aminomethane
ug	microgram

ul ---------- Microliter
uM ---------- Micromolar
XIC ---------- X inactivation center

Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

TRANSCRIPTIONAL REGULATION OF THE HUMAN HYPOXANTHINE PHOSPHORIBOSYLTRANSFERASE GENE BY X CHROMOSOME INACTIVATION

By

Ian Kerst Hornstra

August 1993

Chairperson: Thomas P. Yang
Major Department: Biochemistry and Molecular Biology

Dosage compensation of X-linked genes in male and female mammals is accomplished by random inactivation of one X chromosome in each female somatic cell. As a result, a transcriptionally active allele and a transcriptionally inactive allele of most X-linked genes occupy each female nucleus. To study the mechanism(s) responsible for maintaining this system of differential gene expression, I have examined the 5' region of the human hypoxanthine phosphoribosyltransferase (HPRT) gene on the active and inactive X chromosomes for sequence-specific DNA-protein interactions and DNA methylation. Studies of DNA-protein interactions were carried out in intact cultured cells by *in vivo* footprinting using the ligation-mediated polymerase chain reaction (LMPCR) and dimethylsulfate. Analysis of the

active allele reveals at least six footprinted regions, whereas no specific footprints were detected on the inactive allele. Of the footprints on the active allele, none appear to be specific to X-linked genes, and one appears to define new cis- and trans-acting regulatory elements. Experiments to reconstitute this new DNA-protein interaction in vitro have been performed with crude HeLa nuclear extracts and cloned DNA fragments containing the footprinted region. DNA methylation analysis of the HPRT gene using LMPCR genomic sequencing demonstrates a correlation between transcriptional repression and hypermethylation of the inactive promoter, though complete methylation of the region is not required for inactivation. These results suggest that DNA methylation and/or chromatin structure may have a role in regulating the differential binding of transcription factors to genes on the active and inactive X chromosomes. DNA methylation analysis of the X-linked human FMR1 gene suggests the process of X chromosome inactivation may also be involved in the etiology of the fragile X syndrome. Genomic sequencing of the region within and surrounding the FMR1 trinucleotide repeat indicates all CpG dinucleotides examined are unmethylated in normal and transmitting males, but are methylated in affected males and in a somatic cell hybrid containing the normal inactive X chromosome. Therefore, repression of the FMR1 gene in fragile X males

and silencing of genes on the inactive X chromosome may share common mechanisms.

CHAPTER 1
INTRODUCTION

X Chromosome Inactivation

In placental mammals, the male sex chromosomes have the XY genotype and genes on the male X chromosome are transcriptionally active in somatic tissues throughout development into adulthood. However, female mammals have two X chromosomes and this genotype results in a dosage imbalance of X-linked genes between males and females. To compensate for this dosage imbalance, one X chromosome in each female somatic cell is transcriptionally silenced or inactivated (31,33). This inactivation is developmentally regulated during female embryogenesis (31,33). Initially, both X chromosomes are transcriptionally active in the zygote and remain active until the early blastocyst stage. In the late blastocyst stage, each cell in the embryo proper randomly inactivates either the paternally or maternally derived X chromosome. Once a cell inactivates an X chromosome, the same X chromosome is maintained in the inactive state in all mitotic progeny. Thus, in female cells, a unique system of differential gene expression exists where a transcriptionally active X chromosome and a

transcriptionally inactive X chromosome occupy the same nucleus.

Currently, the process of X inactivation is postulated to occur in three steps (31). Inactivation appears to initiate at a single site on the X chromosome, termed the X inactivation center (31). The inactivation center is hypothesized to be located on the human X chromosome in the Xq13 region. Subsequently, inactivation spreads bi-directionally to inactivate most genes on the entire X chromosome. Interestingly, genes on the short and long arm of the human X chromosome appear to escape inactivation, suggesting these loci lack some type of signal for inactivation (12,23,73) or are otherwise refractory to the inactivation process. However, some genes that escape inactivation in man are inactivated in the mouse. Once inactivation of the chromosome is established, inactivation of the same X chromosome is maintained in all progeny of a given somatic cell. Thus, the pattern of inactivation is transmitted during mitosis and stably maintained within each female somatic cell (121). An exception to the stable maintenance of inactivation is the reactivation of the inactive X chromosome during oogenesis (31-33).

The molecular mechanisms responsible for initiating, spreading, and maintaining X chromosome inactivation are unknown. However, DNA-protein interactions (31,68), chromatin structure (48,80,83), DNA methylation

(44,60,61,74,86,120,126), and DNA replication (30,107) have been postulated to be involved.

Recently, a gene that is expressed exclusively from the inactive X chromosome has been discovered (8). The gene is termed the X inactive specific transcript (XIST). This gene has been localized to the region containing the X inactivation center region on the mouse and human X chromosomes (6,8,11). The mRNA is greater than 15 kb in both man and mouse but lacks a conserved open reading frame (10,46). The RNA is localized almost exclusively to the nucleus and appears to associate with the inactive X chromosome (10). Expression of the XIST gene has been demonstrated to precede X chromosome inactivation, and thus, XIST expression during development may have a role in the initiation of inactivation (46). Despite the exciting information regarding the XIST gene, the mechanism(s) for the initiation, spreading, and maintenance of X chromosome inactivation remain unknown.

The differential expression of genes on the active and inactive X chromosomes is manifested by a difference in nuclease sensitivity of chromatin from the active and inactive alleles of the X-linked (HPRT) and phosphoglycerate kinase (PGK-1) genes (36,59,91,92,122,123). Furthermore, the presence of DNase I hypersensitive sites in the 5' region of the active HPRT and PGK-1 genes (59,92,122,123), and the absence of these hypersensitive sites on the

inactive alleles suggest differential binding of regulatory proteins to genes on the active and inactive X chromosomes (21,34). McBurney (68) has proposed that differential expression of genes on the active and inactive X chromosomes involves specific DNA-binding proteins that bind to cis-acting regulatory sequences near or within the promoter region of each X-linked gene that is subject to inactivation. This hypothesis predicts the existence of a sequence-specific DNA-binding repressor protein that silences genes on the inactive X chromosome, and activator proteins that bind to regulatory regions of genes on the active X chromosome and activate transcription. Recently, in vivo footprint analysis of the human PGK-1 gene has revealed multiple DNA-protein interactions in the 5' region specific to the active allele (83,86); no in vivo footprints were detected on the inactive allele.

In addition to the possible role of DNA-protein interactions and chromatin structure in the maintenance of X chromosome inactivation, DNA methylation has also been implicated. DNA methylation of regulatory regions for some genes has demonstrated a correlation with transcriptional repression (5,57).

In mammals, DNA methylation occurs at the 5 position of cytosines residues in CpG dinucleotides (57). CpG dinucleotides are under-represented in the mammalian genome but occur at high frequency in CpG islands. CpG islands are

about 0.5-2 kb in length, contain a high G+C content, and contain CpG dinucleotides at the frequency expected from base composition (5). CpG islands occur frequently at the 5' end of many constitutive genes and this has been utilized as a marker in searching for genes by positional cloning. Many autosomal CpG islands have been shown to be unmethylated using methyl-sensitive restriction enzymes (5,57,89). However, some 5' GC islands of constitutive X-linked genes on inactive X chromosome are extensively methylated (61,86,120,125). Thus, in general, hypermethylation of gene regulatory regions can be correlated with transcriptional silencing (57) particularly with genes on the inactive X chromosome (4,83).

Many studies have investigated the role of DNA methylation in X chromosome inactivation. There is extensive evidence that strongly supports a correlation between cytosine methylation within 5' CpG islands of constitutively expressed X-linked genes and transcriptional inactivity of genes on the inactive X chromosome (31,68)(see comprehensive reviews). DNA purified from cells containing a inactive X chromosome is not able to transform HPRT- cells to HPRT+ cells, but purified DNA from cells with an active X chromosome is able to transform HPRT- cells to HPRT+ cells (60,112). These experiments suggest that DNA from the inactive X chromosome is physically different from DNA from the active X chromosome. Molecular analysis using

methylation-sensitive restriction enzymes (HpaII, HhaI, etc.) and southern blotting have demonstrated a correlation between hypermethylation of cytosines within the 5' CpG islands and transcriptional silencing in the X-linked human and mouse HPRT genes, human PGK-1 gene, and human G6PD gene (61,86,110,120,126). Furthermore, treatment of hybrid cells containing an inactive X chromosome with a potent demethylating agent, 5-azacytidine (5-azaC), can independently reactivate individual genes on the inactive X chromosome (37,74,110,113). This independent reactivation of individual genes emphasizes that although X chromosome inactivation is a chromosomal-wide process there must be some component of gene regulation at the level of single X-linked genes. Reactivation of the HPRT gene on the inactive X chromosome in a somatic cell hybrid restores the ability to transfect the DNA from HPRT- cells into HPRT+ cells and partially restores the methylation pattern to that of the active X chromosome (using methylation-sensitive restriction enzymes in conjunction with southern blotting). However, a major limitation of methylation analysis using restriction enzymes and southern blotting is that only a fraction of cytosine residues are assayed. Furthermore, methylation analysis of individual restriction sites becomes technically impractical in CpG islands where a high density of restriction sites may be separated by only a few base pairs. To analyze the methylation state of each and every cytosine

residue within the 5' CpG island of the human X-linked PGK-1 gene, Pfeifer et al. (85,86) performed genomic sequencing via ligation-mediated polymerase chain reaction (LMPCR). They found the active PGK-1 allele was completely unmethylated at 120 CpG sites on the active X chromosome, but was essentially completely methylated (118 of 120 CpG sites) on the inactive X chromosome. Therefore, hypermethylation of cytosines within 5' CpG islands of constitutive X-linked genes strongly correlates with transcriptional silencing on the inactive X chromosome.

The mechanism(s) by which DNA methylation inhibits transcription are unknown. Cytosine methylation at cis-acting regulatory elements may interfere with the binding of trans-activating factors (51,117), but some transcription factors like Sp1 and CTF can bind either methylated or unmethylated recognition sequences (3,39,40). In addition, methylated DNA may alter chromatin structure which suppresses transcription (13,49). Another possible mechanism involves the binding of proteins which preferentially bind methylated DNA in a sequence or non-sequence specific manner to inhibit transcription of a methylated promoter (42,58,70,71,116). Recent evidence suggests methylation of sites surrounding the transcription start site (the preinitiation domain) can suppress gene transcription via an indirect mechanism such as a methylated-DNA binding protein (56). Thus, DNA methylation

may suppress the initiation of transcription by a number of mechanisms but the data suggest that methylation within the 5' regulatory region is more important than methylation within the body of the gene.

All the above investigations have studied cells that have already undergone X chromosome inactivation. Studies of the murine HPRT and PGK-1 genes in mouse embryos suggest that DNA methylation in the 5' regions occurs after X chromosome inactivation (62,102). Thus, DNA methylation of X-linked genes on the inactive X chromosome occurs after the initiation of inactivation and appears to stabilize or lock-in the transcriptionally inactive state.

The regulation of gene expression by X chromosome inactivation is likely to be multifactorial involving DNA-protein interactions, chromatin structure, and DNA methylation. Though X chromosome inactivation is a global chromosome-wide process, some degree of regulation at the level of individual X-linked genes must also be involved, as indicated by the ability to independently reactivate individual genes on the inactive X chromosome by treatment with 5-azacytidine (5-azaC) (36,37,74,110,111). In order to investigate transcriptional regulation by X inactivation, it is necessary to analyze individual X-linked genes to provide, if possible, an experimental framework for the global and chromosomal-wide process. The X-linked human HPRT gene has been extensively characterized and on the

inactive X chromosome the HPRT gene is subject to X chromosome inactivation. Similarly, the X-linked human FMR1 gene is inactivated on the inactive X chromosome and the etiology of the fragile X syndrome may involve the process of X chromosome inactivation. Thus, in this dissertation, I have investigated the human hypoxanthine phosphoribosyltransferase and FMR1 genes to provide insight into the mechanism(s) of X chromosome inactivation.

Hypoxanthine Phosphoribosyltransferase

HPRT (E.C.2.4.2.8) catalyzes the salvage of hypoxanthine and guanine to their respective nucleotides, IMP and GMP, by the condensation of 5'-phosphoribosyl-1-pyrophosphate to free hypoxanthine or guanine (104). HPRT is present in all tissues and cells, with elevated levels in the central nervous system, particularly the basal ganglia (104). Complete deficiency of HPRT in man results in the Lesch-Nyhan syndrome and partial deficiency results in hyperuricemia and gout.

The human HPRT gene spans 44 kb and the locus has been entirely sequenced (20). The mRNA is 1.3 kb in length and codes for a protein of 218 amino acids (104). The HPRT gene structure is conserved with 9 exons and the same RNA splicing sites in human and mouse genes (50,82). The mammalian gene is X-linked and constitutively expressed except on the inactive X chromosome, where it is

transcriptionally inactivated. As frequently seen in constitutively expressed genes, the HPRT promoter lacks canonical TATA or CAAT sequences, uses multiple transcription start sites, and is extremely GC-rich (50,82). The human promoter contains four GC box sequences (5'-GGGCGG-3') (50,82), which are potential binding sites for the transcription factor Sp1 (7). Primer extension and nuclease protection studies of the human HPRT promoter region (50,82) have demonstrated multiple sites of transcription in the region from -104 to -169 (relative to the translation start site). In addition, the human promoter is capable of functioning bidirectionally in transient transfection assays when linked to a reported gene (44,93). In transfection studies, the minimal region (-219 to -122) appears to be sufficient for normal levels of HPRT gene expression (93). Furthermore, a putative negative regulatory element has been reported in the region from position -570 to -388.

FMR1 Gene

The fragile X syndrome is the most frequently inherited cause of mental retardation in males with an incidence of about 1 in 2000 (78). This syndrome, characterized by a cytogenetic fragile site at Xq27, is inherited as an X-linked dominant with reduced penetrance and most males who inherit the mutation are affected. The affected males

inherit the fragile X mutation from their mothers. However, the mode of inheritance is unusual because some males possess the fragile X site but are clinically normal. Nonetheless, these clinically normal males, termed transmitting males (101), pass the fragile X mutation to their female progeny. These daughters are phenotypically normal but are obligate carriers of the mutation. Their progeny, who are the grandchildren of the transmitting male, have an increased risk of being clinically affected. Male children have slightly greater than twice the risk of females of being affected. Thus, males who inherit and transmit the genetic lesion do not necessarily manifest the clinical phenotype. Abnormal imprinting of the fragile X chromosome by X chromosome inactivation during female embryogenesis has been postulated to be associated with clinical expression of the fragile X mutation (54).

The recent cloning of the FMR1 gene located at the fragile site on the human X chromosome (52,114) indicates that the fragile X syndrome and the risk of transmitting the disease phenotype are correlated with the size of a $[CGG]_n$ trinucleotide tandem repeat in the 5' untranslated region (26). Normal individuals carry allele sizes between 6 and approximately 50 repeat units that are stable upon transmission. Within fragile X families, two classes of increased and unstable repeat numbers are observed. Transmitting males and most unaffected carrier females carry

a premutation with a repeat number between 50 and approximately 230. Clinically affected individuals exhibit a major expansion of the premutation repeat number to a full mutation with over 230 repeats, often exceeding 1000. The risk for expansion of the premutation to a full mutation increases with the size of the premutation repeat number, and expansion to the full mutation occurs exclusively during female transmission.

However, expansion of the repeat number to the full mutation is apparently not sufficient by itself to produce the disease phenotype. Expression of the disease phenotype appears to be the result of transcriptional repression of FMR1 gene expression (87). This transcriptional silencing is correlated with methylation of a BssHII within the 5' CpG island containing the CGG trinucleotide repeat, a site not methylated in normal or transmitting males (2,79,115). Methylation analysis with additional methyl-sensitive restriction enzymes also indicates hypermethylation of the repeat and its flanking regions (38). Therefore, aberrant methylation at specific sites within the 5' CpG island of the FMR1 gene in affected individuals appears to be correlated with the absence of FMR1 mRNA (and repression of the FMR1 gene) rather than expansion of the repeat number alone. DNA methylation has been widely implicated in gene silencing, particularly in X chromosome inactivation (89). However, the relationship between full expansion of the

repeat,DNA methylation, and repression of the FMR1 gene transcription in fragile X syndrome, as well as the mechanism(s) by which DNA methylation modulates transcription, are unknown.

Specific Aims and Rationale

The goal of this dissertation is to investigate the regulation of transcription by mammalian X chromosome inactivation. Since specific DNA sequences and regulatory proteins appear to have an essential role in many systems of transcriptional regulation, sequence-specific DNA-protein interaction(s) associated with the X-linked human hypoxanthine phosphoribosyltransferase (HPRT) gene have been examined on the active and inactive X chromosomes. Furthermore, the potential role of DNA methylation in modulating transcription has been examined on the human HPRT and FMR1 genes. Identification of differences in DNA-protein interaction(s) between the active and inactive alleles of a single X-linked gene may provide insight into the mechanism(s) for the initiation and establishment of X inactivation at the chromosomal level in the developing female embryo.

The specific aims of this project are briefly discussed below and presented in the following chapters of this dissertation. In Chapter 2, the *in vivo* footprinting of the 5' region of the human HPRT gene on the active and inactive

X chromosomes is presented. The 5' region of the human HPRT gene was studied by in vivo footprinting to identify sequence-specific DNA-protein interactions associated with either the active, inactive, or 5-azacytidine-reactivated allele. In Chapter 3, the in vitro reconstitution of a DNA-protein interaction that is specific to the active HPRT allele is presented. The DNA-protein interaction was identified by the in vivo footprinting studies of the human HPRT gene presented in Chapter 2. Using crude HeLa nuclear extracts and cloned DNA fragments of the HPRT gene containing the in vivo footprint, gel mobility-shift assays have been performed. Chapter 4 contains the DNA methylation analysis of the 5' region of the human HPRT gene on the active and inactive X chromosomes. Cytosine methylation was examined on the active, inactive, and 5-azacytidine-reactivated alleles of the human HPRT gene. The methylation state of specific cytosines has been correlated with transcriptionally activity and with differences observed in the in vivo binding of sequence-specific DNA binding protein(s) between the active and inactive HPRT alleles. Furthermore, in Chapter 5, the high resolution methylation analysis of the human FMR1 gene repeat region in fragile X syndrome is demonstrated. Cytosine methylation within and surrounding the FMR1 trinucleotide repeat was examined in normal males, transmitting males, affected males, and in a somatic cell hybrid containing the normal inactive X

chromosome. The conclusions obtained from these studies and future experimental directions are presented in Chapter 6.

CHAPTER 2
MULTIPLE IN VIVO FOOTPRINTS ARE SPECIFIC TO THE ACTIVE ALLELE OF THE X-LINKED HUMAN HYPOXANTHINE PHOSPHORIBOSYLTRANSFERASE GENE 5'REGION: IMPLICATIONS FOR X CHROMOSOME INACTIVATION

Introduction

The random inactivation of a single X chromosome during normal mammalian female embryogenesis results in a unique system of differential gene expression in which a transcriptionally active X chromosome and transcriptionally inactive X chromosome reside within the same nucleus. The inactivation of genes on one X chromosome in female somatic cells compensates for the dosage imbalance of X-linked genes between the sexes (31,33). The molecular mechanisms responsible for initiating, spreading, and maintaining X chromosome inactivation are unknown. However, DNA-protein interactions (31,68), chromatin structure (48,80,86), DNA replication (30,107), and DNA methylation (47,60,61,74,85,120,126) have all been postulated to be involved. Though X inactivation is a chromosome-wide phenomenon and process, some degree of regulation at the level of individual X-linked genes must also be involved as indicated by the ability to independently reactivate

individual genes on the inactive X chromosome by 5-azacytidine (5-azaC) (36,37,74,110,111).

The differential expression of genes on the active and inactive X chromosomes is manifested by a difference in nuclease sensitivity of chromatin from the active and inactive alleles of the X-linked hypoxanthine-guanine phosphoribosyltransferase (HPRT) and phosphoglycerate kinase (PGK-1) genes (36,59,91,92,122,123). Furthermore, the presence of DNase I hypersensitive sites in the 5' region of the active HPRT and PGK-1 genes (59,92,122,123) and the absence of these hypersensitive sites on the inactive alleles suggest differential binding of regulatory proteins to genes on the active and inactive X chromosomes (21,34). McBurney (68) has proposed that differential expression of genes on the active and inactive X chromosomes involves specific DNA-binding proteins that bind to cis-acting regulatory sequences near or within the promoter region of each X-linked gene that is subject to inactivation. This hypothesis predicts the existence of a sequence-specific DNA-binding repressor protein that silences genes on the inactive X chromosome, and activator proteins that bind to regulatory regions of genes on the active X chromosome and activate transcription. Recently, *in vivo* footprint analysis of the human PGK-1 gene has revealed multiple DNA-protein interactions in the 5' region specific to the active

allele (83,86); no *in vivo* footprints were detected on the inactive allele.

HPRT (EC 2.4.2.8) catalyzes the salvage of hypoxanthine and guanine to their respective nucleotides, IMP and GMP. HPRT is present in all cells and tissues, with elevated mRNA levels and enzymatic activity in the central nervous system, particularly the basal ganglia (104). The mammalian HPRT gene is X-linked and constitutively expressed except on the inactive X chromosome where it is transcriptionally silenced by X chromosome inactivation. As commonly seen in constitutively expressed genes, the HPRT promoter region lacks canonical TATA or CAAT sequences, uses multiple transcription start sites, and is extremely GC-rich with multiple GC box sequences (5'-GGGCGG-3') which are potential binding sites for the transcription factor Sp1 (19,50,82). Primer extension and nuclease protection analyses of the human HPRT promoter region (50,82) have demonstrated multiple sites of transcription initiation in the region from -104 to -169 (relative to the translation start site). Furthermore, the human promoter is capable of functioning bidirectionally *in vitro* (44,93), and a minimal region from -219 to -122 is sufficient for normal levels of HPRT gene expression (93). A putative negative regulatory element has been reported in the region from position -570 to -388 (93).

We now report in vivo footprint analysis of the human HPRT gene 5' region on the active and inactive X chromosomes using the ligation-mediated polymerase chain reaction (LMPCR) (76,85). We demonstrate multiple DNA-protein interactions specific to the active human HPRT allele and the absence of detectable DNA-protein interactions on the inactive allele. One unique footprinted region appears to define a novel regulatory factor(s). These results, in conjunction with similar analysis of the human PGK-1 gene (83,86), have implications for potential models that describe the molecular basis of X chromosome inactivation.

Materials and Methods

Cell Lines

GM00468 (NIGMS Human Genetic Mutant Cell Repository, Camden, NJ) is a normal human 46, XY male fibroblast cell line containing an active X chromosome. Cell line 4.12 (77) (generously provided by Dr. David Ledbetter) is a hamster-human somatic cell hybrid containing only the active human X chromosome in the HPRT-deficient hamster cell line RJK88; RJK88 is a derivative of the V79 Chinese hamster fibroblast cell line and carries a deletion of the endogenous hamster HPRT gene (27). Cell line 8121-6TG D, hereafter referred to as 8121, is a hamster-human somatic cell hybrid containing an inactive human X chromosome in a RJK88 hamster cell background (provided by Dr. David

Ledbetter). The human HPRT gene in 8121 cells was confirmed to be inactive by Northern blot analysis using a human HPRT cDNA probe, by the inability of these cells to grow in HAT-containing medium, by the growth of these cells in the presence of 6-thioguanine, and by the ability to reactivate the human HPRT gene in these cells by 5-azacytidine treatment (see below). HeLa S3 cells were grown in suspension culture and contain at least one active HPRT gene. GM05009b (NIGMS Human Genetic Mutant Cell Repository) is a human 49, XXXXX female fibroblast cell line carrying a single active X chromosome and four inactive X chromosomes (35,109).

In vivo footprint analysis was also carried out on the human HPRT gene of hybrid line 8121 in which the HPRT gene on the inactive human X chromosome was reactivated by treatment with 5-azaC. Cell line 8121R9a is a HPRT reactivant of 8121 grown from a single hypoxanthine/aminopterin/thymidine (HAT)-resistant colony after treatment with 5-azaC essentially as described by Hansen et al. (36). Cell line M22 is a 5-azaC-treated HPRT reactivant of a mouse-human somatic cell hybrid containing an inactive human X chromosome in a murine A9 cell background (this hybrid generously provided by Dr. Barbara Migeon).

All somatic cell hybrids containing an active HPRT gene were cultured using standard techniques in Dulbecco's

modified Eagle's medium (D-MEM) (Gibco) with 10% fetal bovine serum (FBS), 1% penicillin-streptomycin supplement (P-S; Gibco), and supplemented with 1X HAT (0.1 mM hypoxanthine, 0.4 uM aminopterin, 0.016 mM thymidine). Cultures of cell line 8121 were maintained as above without HAT. Human fibroblasts were maintained in Ham's F-12 (Gibco) with 10-20% FBS and 1% P-S. HeLa cells were grown in suspension using suspension modified essential media (S-MEM) with 5% FBS and 1% P-S.

Preparation of DNA--In Vivo Dimethysulfate Treatment and DNA Isolation

Growth media were aspirated from nearly confluent T-150 flasks or 150mm plates, and cells were washed once with 37°C phosphate-buffered saline (PBS). Twenty microliters of dimethylsulfate (DMS) was then added to 20 ml of 37°C PBS (to a final DMS concentration of 0.1%), mixed vigorously, and the final solution gently layered over the cells in each culture flask. Initially, optimal DMS concentration and duration of DMS treatments were empirically determined; all subsequent experiments were carried out using a 5 minute treatment with 0.1% DMS. After treatment with DMS, the DMS-containing PBS was quickly aspirated, and the cells were washed twice with 50 ml of ice-cold PBS. Then 5-10 ml of lysis solution (50 mM Tris, pH-8.5, 50 mM NaCl, 25 mM ethylenediamine tetraacetic acid (EDTA), 0.5% sodium dodecyl sulfate (SDS), 300 ug/ml proteinase K) was added to each

flask or plate and incubated overnight at room temperature. Sodium chloride was added to a final concentration of 200 mM and the lysate was extracted once with phenol, twice with phenol:chloroform:isoamyl alcohol (PCI; 25:24:1), and once with chloroform. DNA in the final aqueous phase was then precipitated with 2 volumes of ethanol and sedimented at 4000 x g for 45 minutes. The supernatant was decanted and the pellet washed with 80% ethanol. After air drying, the pellet was resuspended in either TE (10 mM Tris-HCl, pH-8, 1 mM EDTA) or water.

Occasionally, purified genomic DNA was digested with restriction enzymes (EcoRI or BamHI which do not cut within the region of the human HPRT gene to be analyzed) to reduce viscosity. After restriction enzyme digestion, the DNA was extracted twice with phenol:chloroform:isoamyl alcohol and ethanol precipitated as above. Purified *in vivo* DMS-treated DNA was chemically cleaved at DMS-modified guanine residues using standard Maxam-Gilbert piperidine treatment (67). DNA dissolved in water was first brought to a final concentration of 1M piperidine with a concentrated stock solution of piperidine. DNA in TE was first ethanol precipitated, then redissolved in 1M piperidine. Purified DNA dissolved in 1M piperidine was incubated at 90-95°C for 30 minutes. Samples were then placed on ice, precipitated in 0.3 M sodium acetate (pH-5.2) and 2 volumes of ethanol, and sedimented at 14,000 x g. The resulting pellets were

washed twice with 80% ethanol and dried overnight in a vacuum concentrator. Dried DNA pellets were then resuspended in TE and stored at -20°C. To obtain similar signal intensities among different samples in the final autoradiogram, DNA concentrations were determined spectrophotometrically. In order to confirm that equal amounts of DMS-treated genomic DNA was used in the subsequent LMPCR reactions and that the size distribution of piperidine-cleaved fragments was within the desired size range (average length of 600 bases for *in vivo* DMS-treated samples), a small aliquot of each sample was fractionated on alkaline agarose mini-gels (98) and stained with ethidium bromide.

In Vitro DMS Treatment

Control samples of purified genomic DNA were subjected to Maxam-Gilbert chemical modifications *in vitro* followed by piperidine cleavage. Unmodified genomic DNA was prepared as described above (without prior *in vivo* DMS treatment) and resuspended in water. For each base-specific cleavage reaction, 50 ug of genomic DNA was dried and resuspended in 5 ul of sterile water. In the guanine-specific cleavage reactions, purified genomic DNA was modified with 0.5% DMS for 1 minute at room temperature and processed as described by Maxam and Gilbert (67). Subsequent piperidine cleavage and DNA precipitation were performed as described above.

In order to provide a complete DNA sequencing ladder of the region of interest on each autoradiogram, plasmid pλ4X8-RB1.8 (kindly provided by P. Patel) containing a 1.8 kb EcoRI-BamHI fragment of the human HPRT gene 5' region in pUC8 (82) was linearized with EcoRI, and 2.5 ug of plasmid DNA was chemically modified and cleaved by the standard G, G+A, T+C, C Maxam-Gilbert reactions. The chemically cleaved plasmid DNA was then diluted appropriately to produce autoradiogram signals equivalent to the genomic DNA samples following LMPCR and hybridization with a labelled probe.

Ligation-Mediated PCR

Chemically modified and cleaved DNA was then subjected to amplification by LMPCR essentially as described by Mueller and Wold (76) and Pfeifer et al. (86). The following oligonucleotide primer sets were synthesized (University of Florida Oligonucleotide Synthesis Facility) and used for LMPCR reactions to amplify and analyze specific regions of the human HPRT gene 5' region. For *in vivo* footprint analysis of the lower strand, the following primer sets were used: Set N, primer 1, GATGTGTACCCTGATCTG, and primer 2, GGGTGACTCTAGGACTCTAGGTCTCA; Set A, primer 1, AATGGAAGCCACAGGTAGTG, and primer 2, AGGTCTTGGGAATGGGACGTCTGGT; Set M, primer 1, GAATAGGAGACTGAGTTGGG, and primer 2, GGAGCCTCGGCTTCTTCTGGGAGAA.

For analysis of the upper strand, the following primer sets were used: Set E, primer 1, AGCTGCTCACCACGACG, and primer 2, CCAGGGCTGCGGGTCGCCATAA; Set C, primer 1, AGGCGGAGGCGCAGCAA, and primer 2, GGGAAAGCCGAGAGGTTCGCCTGA; Set R, primer 1, CCAACTCAGTCTCCTATTCA, and primer 2, GAGGGCTCCCTGATTCCCAAACCTA. The region covered by each primer set and the relative positions of the primer sets are diagrammed in Figure 2.1.

After annealing of primer 1 to chemically-cleaved genomic or plasmid DNA, primer extension of the HPRT-specific oligonucleotides using Sequenase (US Biochemicals) was performed as described by Pfeifer et al. (86) except that 7-deaza-dGTP was substituted in a 3:1 molar ratio with dGTP. 5 ug of chemically cleaved genomic DNA was used for each Sequenase reaction. Following extension of primer 1 by Sequenase, blunt-end ligation of the asymmetric double-stranded linker was performed as described by Mueller and Wold (76). Ligated DNA was ethanol precipitated in 2.5 M ammonium acetate and redissolved in 20 ul sterile water. The appropriate region of the human HPRT gene was then amplified by the polymerase chain reaction (PCR) with Taq DNA polymerase (Perkin-Elmer Cetus) using primer 2 from each primer set and the longer oligonucleotide of the asymmetric linker as primers (76). Again, 7-deaza-dGTP was substituted for dGTP in a 3:1 molar ratio with dGTP to allow the amplification of regions with extremely high G+C content.

After 18 cycles of PCR (using a Coy Tempcycler), the DNA was extracted once with PCI, once with chloroform, and precipitated with ammonium acetate and ethanol as before. The resulting pellet was washed with 1 ml of 80% ethanol, dried in a vacuum concentrator, resupended in 20 ul water, and stored at -20°C. Each of the HPRT-specific primer sets was used individually for LMPCR because multiplex analysis (83,86) using two or more primer sets in each LMPCR reaction occasionally yielded artifacts or variability between experiments.

Gel Electrophoresis and Electrotransfer

Five microliters of each PCR reaction was dried and resuspended in 2.0 ul of formamide-dye solution (98% formamide, 0.25% xylene cyanol, 0.25% bromophenol blue, 10 mM EDTA, pH-8). The redissolved samples were denatured at 95°C for 5 minutes, and quenched on ice. Denatured samples were then loaded onto a 0.04 cm thick, 8.3 M urea, 6% polyacrylamide (29:1 acrylamide:bis-acrylamide) DNA sequencing gel in 1 X TBE (50 mM Tris, 50 mM Boric acid, 2 mM EDTA, pH 8.3). Following electrophoresis at 40-50°C, the gel was transferred to Whatman 541 SFC paper. DNA in the gel was then electrotransferred to Hybond N+ nylon membrane (Amersham) using an electroblotting apparatus (Polytech Products, MA) at 110 volts, 2 amperes, in transfer buffer (40 mM Tris, 40mM boric acid, 1.6 mM EDTA, pH 8.3) for 45

minutes as described by Church and Gilbert (15). After transfer, the nylon membrane was rinsed briefly in transfer buffer and then dried in a vacuum oven at 80°C for 1 hour.

Probe Synthesis, Hybridization, and Washing

The ^{32}P-labelled hybridization probes used to visualize the DNA sequencing ladder and *in vivo* footprints were synthesized from a single-stranded M13 template using a modification of the procedure described by Church and Gilbert (15). To generate the appropriate single-stranded HPRT-specific templates for probe synthesis, the 1.8 kb EcoRI-BamHI human HPRT 5' genomic fragment of plasmid pλ4X8-RB1.8 (82) was cloned into the EcoRI-BamHI sites of both M13mp18 and M13mp19, yielding two subclones with the insert in different orientations, with each single-strand template carrying a different strand of the human HPRT gene 5' region. Large-scale preparations of each single-stranded M13 template DNA was performed as described by Sambrook et al. (98).

Synthesis of the labelled single-stranded hybridization probe from the appropriate M13 template was similar to that described by Church and Gilbert with one notable exception. Synthesis of the labelled probe was primed using primer 2 from the appropriate HPRT-specific LMPCR primer set rather than priming with the M13 universal primer. This modified procedure for probe synthesis was performed as follows. One

half picomole of the appropriate purified M13 template (containing one strand of the human HPRT 5' region), 5 ul of a 1 pmol/ul solution of the appropriate primer 2 (which is complementary to the M13 template), and 2.5 ul 10X Klenow buffer (10X buffer: 2 M NaCl, 500 mM, Tris pH-8) were combined in a 1.5 ml microcentrifuge tube. The mixture was denatured at 95°C for 5 minutes and then incubated at 50°C for 30 minutes. Following annealing, 5 ul of 50 mM MgCl2, 5 ul 0.1M dithiothreitol (DTT), 2 ul of a 3 mM solution each of dATP, dGTP, and dTTP, 10 ul dCTP-α32P (Amersham, 3000 Ci/mmol, 10 uCi/ul), 2 ul Klenow fragment (5 U/ul) (Ambion) were added and incubated at 37°C for 45 minutes. Then, 120 ul of formamide-dye solution was added, the mixture denatured at 95°C for 10 minutes, quenched on ice, and loaded onto a 1.5 mm-thick 6% denaturing polyacrylamide gel (6% acrylamide, 40:1 acrlyamide:bis-acrylamide, 8.3 M urea) in 2 x TBE (100 mM Tris, 100 mM boric acid, 4 mM EDTA). Electrophoresis was continued until the xylene cyanol and bromophenol blue markers were separated by 4-5 cm, then labelled probe was excised from the gel. The optimal probe length is just above the xylene cyanol dye, though shorter and longer probes have been used with success. The probe length is controlled by adjusting the ratio of template DNA to radiolabelled dCTP. The portion of the acrylamide gel containing the probe was cut from the remainder of the gel with a razor blade, crushed into a fine paste with a glass

rod, and suspended in 4-6 ml of hybridization solution (0.25 M Na_2HPO_4 brought to pH 7.2 with phosphoric acid, 7% SDS, 1% fraction V bovine serum albumin (Sigma), 1 mM EDTA, as described by Church and Gilbert (15) at 65°C. Simultaneously, the nylon blot was prehybridized for 10-15 minutes with 15 ml of hybridization solution at 65°C in the glass tube of a hybridization chamber (Robbins Scientific, CA). After 15 minutes, the prehybridization solution was discarded and the slurry containing the labelled probe was added directly to the hybridization tube. The blot was hybridized for 6-8 hours at 68°C, the hybridization solution discarded, and the blot quickly and vigorously rinsed 3-4 times with 50-100 ml of wash solution (40 mM Na_2HPO_4 brought to pH 7.2 with phosphoric acid, 1% SDS, 1 mM EDTA as described by Church and Gilbert) at 65°C in the hybridization tube. The blot was transferred to a shaking water bath (Bellco) containing wash solution at 65°C and the wash solution was exchanged every 10-15 minutes until non-specific background was removed. The blot was then covered with plastic wrap and exposed to either Kodak X-OMAT AR film or Amersham Hyperfilm MP without intensifying screens for 3 hours to several days.

Results

The 5' region of the human HPRT gene on the active and inactive X chromosomes was examined in vivo for sequence-specific DNA-protein interactions. The region spanning positions -530 to -14 (relative to the translation initiation codon) was subjected to in vivo footprint analysis using a modification of the ligation-mediated PCR technique described by Mueller and Wold (76) and Pfeifer et al. (83).

This analysis was performed on seven different cell lines to examine the in vivo footprint pattern of either the active or the inactive HPRT allele. Hybrid cell line 4.12 contains only the active human X chromosome in hamster cell line RJK88 which carries a deletion of the hamster HPRT gene (21). Thus, any in vivo footprint detected on the HPRT gene will be specific to the active human HPRT allele. Similarly, cell line 8121 is a human-hamster hybrid that contains the inactive human X chromosome in a RJK88 hamster cell background. Footprints detected on the HPRT gene in this cell line will be associated with the inactive human allele. Since sequence-specific DNA binding proteins in cell lines 4.12 and 8121 will most likely be of hamster origin and will be bound to heterologous human HPRT DNA sequences, normal human male fibroblasts and HeLa cells were

included in the analysis as controls. Both of these cell lines carry an active human HPRT gene interacting with endogenous human DNA-binding proteins, and were useful for identifying footprints that may have been due to artifacts of a heterologous human-hamster hybrid system. All footprints observed in hybrid 4.12 were also present and identical in the male fibroblast cell line and HeLa cells (see below). To confirm that in vivo footprints on the inactive HPRT allele in hybrid 8121 are also present in intact female human cells, a human fibroblast cell line carrying 5 X chromosomes (karyotye 49, XXXXX) was also analyzed. Because this cell line carries 4 inactive human X chromosomes and a single active X chromosome (35,109), the predominant in vivo footprint pattern from the human HPRT gene will be derived from the inactive allele. Therefore, analysis of these cells will confirm results from hybrid cell line 8121 (carrying the inactive X chromosome).

In addition to the in vivo footprint pattern on the active and inactive X chromosomes, the footprint pattern of 5-azacytidine-reactivated HPRT genes on the inactive X chromosome was examined. Cultures of 8121 cells (carrying an inactive X chromosome) were plated at low density, grown in the presence of 5-azaC, and selected for reactivation of the human HPRT gene in HAT-containing medium. Cells that carried a reactivated HPRT gene were HAT-resistant and isolated as single cell-derived colonies. Twelve HAT-

resistant colonies were isolated and subjected to Northern blot analysis to determine the relative level of HPRT mRNA in each isolate (data not shown). The isolate that displayed the highest level of HPRT mRNA (cell line 8121R9a) was used for in vivo footprint analysis. In vivo footprint analysis was also performed on cell line M22, a 5-azaC-reactivated human HPRT gene in a HPRT-deficient mouse A9 cell background (120).

Figure 2.1 shows the relative location of the oligonucleotide primer sets used for LMPCR in vivo footprinting of the 5' region of the human HPRT gene. The region from positions -530 to -14 was analyzed for sequence-specific DNA-protein interactions on both strands. More extended analysis of the lower strand of the region spanning -13 to +42, and the upper strand of the region spanning -531 to -580, was also possible using primer sets M and R, respectively.

Results of LMPCR in vivo footprinting of the upper strand in the region of the multiple transcription start sites (50,82) using primer set E is shown in Figure 2.2. A single guanine showing strong enhanced reactivity to DMS is detected at position -91 in all samples prepared from cells treated in vivo with DMS that carry an active X chromosome or a 5-azaC-reactivated human HPRT gene. This enhanced cleavage site is not detected in purified DNA samples (from the same cell lines) that were treated with DMS after DNA

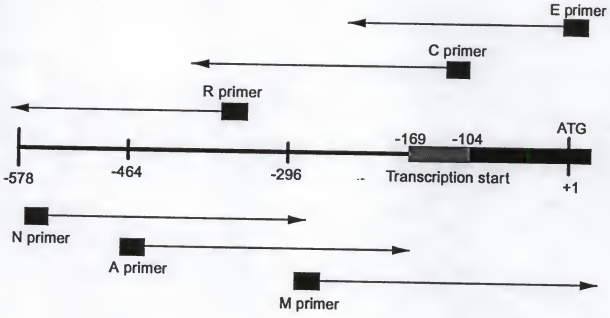

Figure 2.1 - Location of primers used in the LMPCR analysis of the human HPRT 5' region. The numbered line represents the human HPRT gene 5' region with positions numbered relative to the translation initiation codon. The large rectangle represents the first exon with the cross-hatched portion signifying the region of multiple transcription initiation sites (50,82). The smaller rectangles above and below the numbered line indicate positions of the PCR primer sets used in the LMPCR footprinting analysis. Primer sets N, A, M, are complimentary to the lower strand sequence and primers E, C, R, are complimentary to the upper strand sequence. Lines with arrowheads indicate the region and direction resolved by each primer set.

isolation, nor is it detected in the in vivo-treated sample of cell line 8121 which contains the inactive human X chromosome. Very weak protection from DMS is also observedat the guanine residue at position -93. These features are the only evidence for a footprint on the upper strand between positions -14 and -162, and all samples with an active human HPRT gene display the identical footprint pattern. This includes samples where the human HPRT gene is active in human, hamster, and mouse cell backgrounds as well as reactivated with 5-azaC. Interestingly, a palindrome of the sequence GCGGC, with a dyad axis of symmetry between positions -92 and -91, includes both the site of strong enhanced DMS reactivity and the weakly protected guanine residue. However, because this footprint is not detected in purified DNA treated with DMS (in vitro treated samples), it is very likely that the footprint is due to binding of a protein in vivo rather than secondary structure in purified DNA. Due to the strength of this enhanced DMS reactivity at position -91, the sample from the 49, XXXXX human fibroblast cell line also shows a readily detectable signal despite the presence of only a single active HPRT allele among five HPRT genes (four of which are inactive).

Analysis of this same region on the opposite strand (lower strand) was carried out using PCR primer set M; the results are shown in Figure 2.3. Comparison of the cleavage patterns and relative band intensities between DMS-treated

purified DNA samples and *in vivo* DMS-treated samples, reveals two enhanced DMS-reactive sites, one at position -75, and another single enhancement at position -90. As with the footprint in this region on the upper strand, these enhanced cleavages occur only in samples where intact cells carrying an active human X chromosome or active human HPRT gene were treated *in vivo* with DMS prior to DNA purification. One site of enhanced reactivity (at position -90) occurs within the immediate region of the footprint observed on the opposite (upper) strand (at the strong enhancement at position -91). The enhancement at position -75 on the lower strand is 16 nucleotides downstream of the other protection/enhancements in this region, and it is unclear if this single enhancement represents a separate footprint (i.e., different DNA binding protein) or is part of the DNA-protein interaction occurring around position -91. The DNA sequence containing the -91 footprint has not been reported to be a site for binding of a transcription factor (24,63).

The -91 footprint is unusual because it consists of three sites of enhanced DMS reactivity with no guanine nucleotides showing strong protection from DMS. It is possible that the DNA-binding protein(s) interacting at this site does not maintain close contacts with guanine residues

Figure 2.2 - In vivo footprints in the region spanning positions -75 to -98 using primer set E. This autoradiogram shows the guanine-specific cleavages and sequencing ladder from the upper strand. The nucleotide sequence in the region of each footprint and the position of each nucleotide relative to the translation initiation codon is shown to the left of each sequencing ladder. Open circles to the right of the nucleotide sequence represent the sites of enhanced DMS reactivity, and solid circles represent sites of protected guanine nucleotides. For the gel lane designations, DNA denotes purified naked DNA isolated from the appropriate cell line and treated with DMS in vitro. Cells denotes samples that were obtained from intact cells treated in vivo with DMS. Xa indicates samples containing the active human X chromosome, Xi indicates samples containing the inactive human X chromosome, and Xr and 5-AzaC indicate samples from rodent-human hybrid cell lines containing a 5-azacytidine-reactivated human HPRT gene on the inactive X chromosome in either a hamster (lane H; cell line 8121R9a) or mouse (lane M; cell line M22) cell background. XY denotes samples prepared from normal diploid male human fibroblasts (cell line GM00468). Hybrid denotes samples prepared from hamster-human somatic cell hybrids containing either the active (cell line 4.12) or inactive (cell line 8121) human X chromosome. HeLa denotes HeLa cells, and Xa/4Xi denotes samples from a 49, XXXXX female fibroblast cell line (cell line GM05009b).

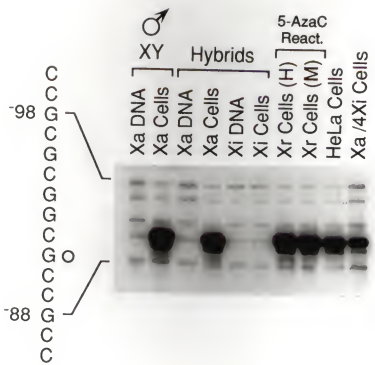

Figure 2.2 - In vivo footprints in the region spanning positions -75 to -98 using primer set E.

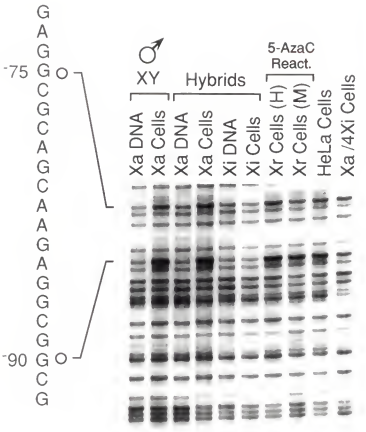

Figure 2.3 - In vivo footprints in the region spanning positions -75 to -98 using primer set M. This autoradiogram shows the guanine-specific cleavages and sequencing ladder from the lower strand. Lane designations and symbols are identical to those in Figure 2.2.

within the binding site; however, near the edge of the footprinted region, the three footprinted guanine residues (at positions -75, -90, and -91) may be more accessible to DMS and therefore react more frequently. To verify the presence of a DNA-protein interaction in this region, in vitro gel mobility-shift assays have been performed; a labelled DNA fragment carrying this footprinted region (and excludes other regions that exhibit in vivo footprints) displays multiple retarded bands when incubated with a crude HeLa cell nuclear extract in the presence of specific and non-specific competitor DNA (I.K. Hornstra and T.P. Yang, unpublished data).

Proceeding upstream from position -91, no evidence for footprints on either strand is detected in any of the samples until position -159 is analyzed with primer set M on the lower strand. In all samples carrying an active human HPRT gene that were DMS-treated in vivo, the guanine nucleotide at position -159 shows enhanced DMS reactivity followed by protected guanines at positions -160 and -165 (Figure 2.4). Again, no evidence for a corresponding footprint is detected in vivo-treated samples from the somatic cell hybrid 8121 containing the inactive X chromosome. Similarly, the cleavage pattern of the 49, XXXXX sample was comparable to the pattern seen with both naked DNA and hybrid 8121. Further evidence for a footprint in this region from samples containing an active HPRT gene

is detected on the upper strand using primer set C. As shown in Figure 2.5, enhanced DMS reactivity at the guanine residue in position -163 is followed by 4 protected guanine residues (positions -164 to -168). Weaker (but significant) protection is observed in the 5-azaC reactivated human HPRT gene in the mouse cell background (cell line M22); this appears to be true for nearly all of the footprints detected in this cell line, and the reason for this is unclear. This footprinted region (from position -159 to -168) contains a canonical GC box (GGGCGG; designated GC box I in Figures 2.4 and 2.5) suggestive of binding in vivo of the transcription factor Sp1 (7,19)--or a rodent homologue of Sp1--to the active human HPRT allele and in 5-azaC reactivated HPRT genes.

The in vivo footprint associated with GC box I on the active HPRT gene is followed in these same samples by a series of DMS protected sites and enhanced reactivity sites immediately upstream at guanines in three additional GC box sequences (designated GC boxes II, III, and IV) using primer set C. As seen in Figures 2.4 and 2.5, in vivo footprints are detected on both strands between positions -172 to -190 (that includes GC box II), -194 to -205 (that includes GC box III), and -207 to -215 (that includes GC box IV). Each of these in vivo footprints is detected only in samples containing an active or reactivated human HPRT gene.

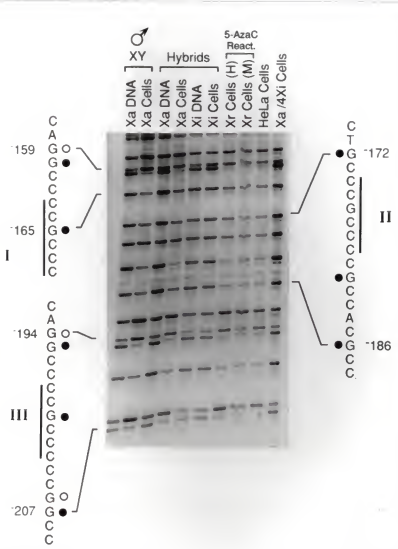

Figure 2.4 - In vivo footprint analysis of the region spanning positions -159 to -215 using primer set M. The autoradiogram shows the guanine sequencing ladder of the lower strand. Lane designations and symbols are identical to those in Figure 2.2. Solid vertical lines indicate the position of GC boxes, and roman numerals adjacent to GC boxes correspond to positions of GC boxes indicated in Figure 2.7 and discussed in text.

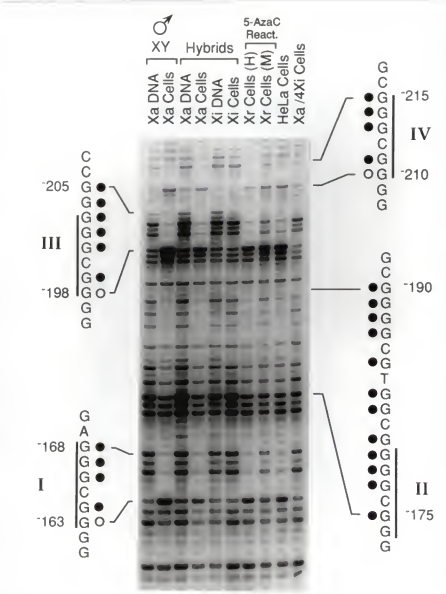

Figure 2.5 - In vivo footprint analysis of the region spanning positions -159 to -215 using primer set C. The autoradiogram shows the guanine-specific sequencing ladder of the upper strand. Lane designations and symbols are identical to those in Figure 2.2. Solid vertical lines indicate the position of GC boxes, and roman numerals adjacent to GC boxes correspond to positions of GC boxes indicated in Figure 2.7 and discussed in text.

However, only the sequence surrounding GC box III (GGGGCGGGGC) conforms to the consensus Sp1 binding sequence described by Briggs and Tjian (7). In addition to the potential binding of Sp1 at each of the four GC boxes, another potential Sp1 binding sequence (GGGGCGTGGC;1) immediately upstream of GC box II (from position -181 to -190) is also included within a footprinted region on the active HPRT gene, though it does not carry a classical GC box sequence. Thus, the active (and 5-azaC-reactivated) human HPRT promoter region exhibits in vivo footprints over 5 potential Sp1 binding sites. Interestingly, the region surrounding the footprint between positions -175 and -190 contains a direct repeat of the sequence GCGGGGCG.

Further upstream from the multiple footprints associated with GC boxes I-IV, primer set A detects a series of three protected guanine residues on the active HPRT alleles between positions -265 and -267 on the lower strand (see Figure 2.6), though the degree of protection appears to vary according to the cell line analyzed. The footprint is readily detected in diploid male human fibroblasts, hybrid cell line 4.12 containing the active human X chromosome, and a 5-azaC reactivated human HPRT gene in a hamster-human hybrid (cell line 8121R9a), while clearly not present in hamster-human hybrid 8121 carrying the inactive human X

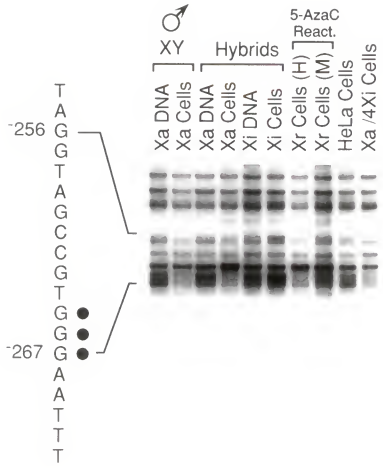

Figure 2.6 - In vivo footprint analysis of the region spanning positions -256 to -267 using primer set A. The autoradiogram shows the guanine sequencing ladder of the lower strand. Lane designations and symbols are identical to those in Figure 2.2.

chromosome. However, the three guanine residues are only weakly protected, if at all, in two other cell lines containing an active human HPRT gene, the 5-azaC reactivated HPRT gene in a mouse-human hybrid (cell line M22), and HeLa cells. The basis of the weak protection of this region in HeLa cells is unknown, particularly since HeLa cells show strong footprints at all of the other footprinted regions, and a factor binding to this DNA sequence (5'-TGGGAATT-3') has been reported in HeLa cells (43); see Discussion below). The reason for very weak protection at this position in the mouse-human hybrid reactivant is also unknown. However, this cell line also shows slightly weaker protections in the region of the GC boxes (see Figs. 2.4 and 2.5), perhaps suggesting that some mouse binding factors may not interact identically with binding sites in human DNA compared to the homologous factors in man and hamster. No footprint of this region is observed in any cell line on the upper strand using primer set C, perhaps because this region on the upper strand is deficient in guanine residues. Curiously, unlike all of the other footprints observed in this study, this region does appear to demonstrate full protection in the 49, XXXXX human fibroblast cell carrying 4 inactive X chromosomes (Figure 2.6, lane Xa/4Xi), suggesting that this region may be bound by a protein on most or all of the multiple inactive X chromosomes as well as the active X chromosome.

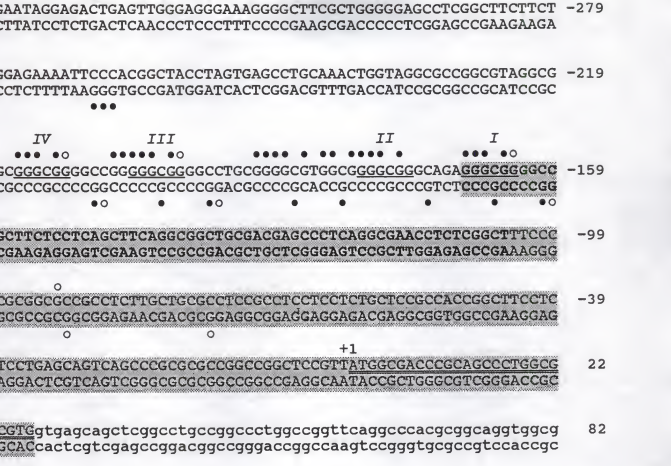

Figure 2.7 - Summary of in vivo footprint analysis of the human HPRT gene 5' region. The sequence of the human HPRT 5' region indicating positions of in vivo footprints on the active HPRT allele. Numbering of nucleotides begins with +1 at the translation initiation codon. The shaded region indicates the first exon. The nucleotides shown in bold-face within the first exon represent the region of multiple transcription initiation sites (50,82). The double underlined region denotes the protein coding region within exon 1. The region shown in lower case letters indicates nucleotides within the first intron. The regions underlined with a single line indicate the positions of GC boxes. Each of the 4 GC boxes is numbered with a roman numeral that corresponds to the roman numerals indicating GC boxes in Figures 2.4 and 2.5. Closed circles indicate the position of protected guanine residues, and open circles indicate the position of enhanced DMS reactivity. Circles above the nucleotide sequence indicate footprints detected on the upper strand and circles below the sequence indicate footprints detected in the lower strand.

No evidence for any other footprints in the region from [-580 to +42] is detected on either strand. This includes the region from [-570 to -388] that is reported to contain a negative regulatory element by deletion analysis (93). Figure 2.7 shows the nucleotide sequence of the 5' region from the human HPRT gene and summarizes the DMS *in vivo* footprint data by indicating the position of all DMS protected sites and sites of enhanced DMS reactivity detected in this study.

Discussion

In vivo DMS footprint analysis of the immediate 5' region of the human HPRT gene in a variety of cell lines carrying active and/or inactive human X chromosomes has revealed multiple footprints specifically on the transcriptionally active allele. At least six *in vivo* footprints are located on the active, or 5-azaC-reactivated, HPRT gene and are presumed to indicate sites of sequence-specific DNA-protein interactions. The footprint patterns in cell lines carrying an active human HPRT gene are identical despite differences in the species of the background cell line (human, hamster, or mouse), suggesting the DNA-binding proteins from the rodent species are interacting with the human HPRT DNA sequences in a manner identical to the human binding proteins seen in normal human male cells. The appearance of these footprints correlate

with transcriptional activity of the human HPRT gene and the presence of a nuclease hypersensitive site in the 5' region of the transcriptionally active gene (23,55). In contrast, the HPRT gene on the inactive X chromosome--with a single apparent exception in the 49, 5X female cell line (see below)--appears to be devoid of detectable sequence-specific in vivo footprints. Furthermore, the DMS reactivity patterns of the inactive HPRT gene in hybrid 8121 is essentially indistinguishable from that of naked DNA.

DNA-Protein Interactions Specific to the Active HPRT Allele

The DNA sequences associated with each of the in vivo footprints on the active HPRT gene include sequences previously identified as binding sites for regulatory proteins as well as DNA sequences not previously reported to be target sites for DNA-binding proteins. The DNA sequence contained within (or immediately adjacent to) the footprint associated with the strong DMS-reactive site at position -91 on the upper strand and enhancements at -90 and -75 on the lower strand (termed the -91 footprint) appears to represent a new cis-acting regulatory element and a target sequence for a new DNA-binding protein(s). A DNA data search using the DNA sequence from the immediate region containing the enhanced DMS-reactive sites at position -91 to position -75 did not yield clear sequence identity with any previously described regulatory elements among vertebrate control DNA

sequences (24,63). The position of this footprinted region just 3' to the multiple sites of transcription initiation (-104 to -169) suggests the protein(s) associated with this DNA sequence may function in transcription initiation as has been postulated for other DNA-binding regulatory factors located in a similar position. These factors include HIP-1 (69), Inr (103), YY1 (100), and TFII-I (97). Comparison of the DNA sequence in the -91 footprint with the DNA sequences bound by these initiation factors yielded no significant sequence similarity between these cis-acting elements and the -91 footprint. This suggests that the DNA-protein interaction(s) in the -91 footprint may represent a new regulatory element involved in transcription initiation. Notably, the DNA sequence within the -91 footprint region does not bear significant homology to the binding site of HIP-1, a factor associated with transcription initiation of the dihydrofolate reductase (DHFR) gene, a constitutively expressed gene with a promoter structure similar to that of HPRT. Furthermore, no evidence for in vivo binding of HIP-1 was detected in the human HPRT 5' region.

Recently, Rincon-Limas et al. (93) have reported that promoter DNA sequences between -219 to -122 are necessary and sufficient for normal expression levels of the human HPRT gene by DNA transfection and transient expression assays. However, the region spanning this promoter fragment does not include the region carrying the -91 in vivo

footprint. Thus, the DNA-protein interaction(s) represented by the -91 footprint does not appear to be required for normal function of the human HPRT promoter by this assay. Assuming the -91 in vivo footprint does represent a functional sequence-specific DNA-protein interaction, two interpretations of these data are possible. Either the DNA-protein interaction represented by the -91 in vivo footprint is not directly involved in activation of transcription and serves another function in HPRT gene expression, or transient expression assays do not accurately duplicate expression and regulation of the intact HPRT gene in vivo. More recent studies of the -219 to -122 promoter fragment in transgenic mice indicate additional DNA sequences from the HPRT gene 5' region are required for normal promoter function (F. Rincon-Limas and P. Patel, personal communication).

Upstream of the -91 footprint, a closely spaced cluster of at least four in vivo footprints are observed between positions -159 to -215 in the HPRT gene on active human X chromosomes and on 5-azaC-reactivated HPRT genes. These footprints are not seen in somatic cell hybrid 8121 carrying the inactive human X chromosome or in the 49, XXXXX human fibroblasts cells that contain 4 inactive X chromosomes. The close proximity of the footprints in this region makes it difficult to infer the actual number of discrete binding sites for regulatory proteins. However, this region

contains four copies of the hexanucleotide sequence 5'-GGGCGG-3', each of which is included in regions that exhibit an in vivo footprint on the active human HPRT gene. This sequence, termed a GC box, is the core sequence of the binding site for the transcription factor Sp1 (7), suggesting a role for at least 4 Sp1 molecules (and its rodent homologues in somatic cell hybrids) in transcription of the human HPRT gene in vivo. However, these in vivo footprinting studies do not permit identification of the proteins bound at each of the footprinted sites, and it is possible that a protein(s) other than Sp1 may be interacting at these apparent Sp1 binding sites. Nonetheless, the footprints associated with three of the four GC boxes (GC boxes I, III, and IV) exhibit a very similar pattern of DMS protection and enhanced reactivity in vivo suggesting that the same protein(s) may be bound in vivo at these three sites. The in vivo footprint that includes GC box II (from positions -172 to -190) is larger and displays a slightly different pattern of DMS protection and enhanced reactivity (for example, lack of sites showing enhanced reactivity) from GC boxes I, III, and IV. Closer examination of the DNA sequence in this region reveals another potential Sp1 binding site immediately upstream of GC box II (between positions -181 to -190) that does not contain a classical GC box. Slight DNA sequence variations in each of these 5 potential Sp1 binding sites may account for the slight

difference in vivo footprint patterns associated with each site. Furthermore, only GC box III and the potential Sp1 binding site upstream of GC box II match the reported consensus binding site for Sp1 (7). Thus, the DNA sequences containing GC boxes I, II, and IV may represent additional degeneracy in the binding site sequence for Sp1 (or binding of a protein(s) other than Sp1).

Further upstream of the GC boxes in a region from position -265 to -267, three adjacent guanine nucleotides exhibit some degree of protection from DMS in vivo in all cell lines carrying an active human HPRT gene. The DNA sequence including and surrounding the protected guanine residues contains a potential binding site for the transcription factor AP-2 (118), as well as factors E2aE-CB and E4F2, cell-encoded factors that bind to this sequence in the adenovirus E2A and E4 genes, respectively (43,63). This in vivo footprinted region in the human HPRT gene is also not included within the minimal promoter fragment (from -219 to -122) previously identified as having full promoter function in transient expression assays (93). Curiously, the presence of this in vivo footprint does not appear to completely correlate with transcriptional activity of the human HPRT gene (see Results above; Fig 2.6). Furthermore, the 49, XXXXX human female cells carrying a single active X chromosome and four inactive X chromosomes appears to display full protection in this region. This would suggest

that this factor is bound to most, if not all, of the HPRT gene copies in this cell line, regardless of whether they are on the active or inactive X chromosomes. The role of this factor in the differential expression of the HPRT gene on the active and inactive X chromosomes is unclear.

No other in vivo footprints in the immediate 5' region on either the active or inactive human HPRT alleles are detected in this study. This includes the region from -570 to -388 reported to contain a negative regulatory element (93). However, DMS footprinting only reveals very close contacts between DNA-binding proteins and guanine residues. Therefore, DNA-binding proteins that are weakly associated with guanine residues, or that bind DNA sequences lacking guanines, may not be detected by DMS footprinting. However, it is possible that in vivo footprint analysis using DNase I (83) may reveal DNA-protein interactions not readily detectable by DMS footprinting.

Comparison of in Vivo Footprinting of Human HPRT and PGK-1

In vivo footprint analysis of the human HPRT gene now permits a comparison with similar studies of the human PGK-1 gene on the active and inactive X chromosomes by Pfeifer et al. (83,86) to identify a common basis for the differential expression of these genes on the active and inactive X chromosomes. These studies reveal both significant similarities and differences. The promoter regions of both

genes are GC-rich, lack TATA boxes, and display multiple *in vivo* footprints only on the active X chromosome and 5-azaC-reactivated genes. The promoter region of both genes on the active X chromosome also exhibits *in vivo* footprints associated with multiple GC boxes, suggesting the ubiquitous transcription factor Sp1 is involved in the transcriptional activation of both genes. No *in vivo* footprints are detected using DMS on the inactive HPRT allele (with one possible exception in 49, XXXXX cells; see above) or with DMS and DNase I on the inactive PGK-1 allele (83,86). Thus, in both genes, no sequence-specific DNA-protein interaction is present on the inactive allele in all cells carrying an inactive X chromosome.

Other than the presumptive Sp1 *in vivo* footprints associated with the multiple GC boxes and/or Sp1 consensus sequences in each gene, no DNA sequences common to both genes are footprinted. For instance, the human PGK-1 gene does not display a footprint in the region equivalent to the -91 footprint region in human HPRT (just downstream of the multiple transcription start sites in both genes). Thus, there appears to be no novel DNA-binding regulatory factor or DNA-protein interaction that is specific for X-linked genes (or even to X-linked housekeeping genes) either on the active or inactive X chromosomes.

Implications for X Chromosome Inactivation

In vivo footprinting studies of the X-linked human HPRT and PGK-1 (83,86) genes provide insight into potential mechanisms associated with this unique system of coordinately regulated differential gene expression. First, these studies do not appear to support the hypothesis that X inactivation is a process regulated by a specific DNA sequence that binds either activator or repressor proteins within the promoter region of each X-linked gene subject to inactivation (68). The absence of an *in vivo* footprint on the inactive allele of the HPRT and PGK-1 genes argues against a sequence-specific repressor protein binding to each X-linked gene subject to X inactivation which silences genes on the inactive X chromosome. These data also argue against models for X inactivation that require a unique activator protein(s) that specifically potentiates transcription of X-linked genes (on the active X chromosome) since a novel *in vivo* footprinted DNA sequence common to both HPRT and PGK-1 has not been identified on the active allele of both genes. However, it is possible that the binding sites for important regulatory proteins may be located further upstream of the gene, within the body of the gene, or further 3' of the gene, rather than in the immediate 5' region analyzed in these studies.

A role for DNA methylation in X inactivation has been suggested, in part, by the relative hypermethylation of

cytosine residues in the GC-rich island in the 5' region of X-linked housekeeping genes on the inactive allele compared to the active allele (47,85,86,120). Meehan et al. (72) and Huang et al. (42) have described DNA-binding proteins that preferentially bind to methylated DNA. These proteins could potentially play a role in silencing transcription of housekeeping genes by specifically binding to hypermethylated GC-rich promoter regions (or GC islands) on the inactive X chromosome. No evidence for such proteins has been detected in the 5' region of either the HPRT or PGK-1 (83,86) genes by in vivo footprinting of the inactive alleles. However, it is still possible that these proteins may be present on the inactive X chromosome and are not detected by these studies due to lack of DNA sequence specificity or weak binding (83,86).

The presence of multiple footprints on the active X chromosome, and the lack of footprints on the inactive X chromosome, suggests that transcription factors in female nuclei--while able to bind and activate transcription of genes on the active X chromosome in the same nucleus--may be unable to gain access to their target DNA sequences on the inactive X chromosome, or are unable to form stable sequence-specific DNA-protein complexes on the inactive X chromosome. One possibility for preventing binding of factors on the inactive allele of X-linked genes is that DNA methylation may interfere directly with formation of stable

sequence-specific DNA-protein complexes (51,117). However, this may not be a general mechanism for preventing stable binding of transcription factors to the inactive X chromosome because binding of at least one potential factor identified by *in vivo* footprinting on the active X chromosome - Sp1 - is not affected by methylation within its binding site when assayed *in vitro* (39,40). An alternative mechanism for the differential binding of transcription factors to the active and inactive alleles of X-linked genes may involve chromatin structure. The presence of nucleosomes at DNA binding sites (83) or higher order chromatin structure on the inactive X chromosome may prevent binding of transcription factors to their binding sites, while the chromatin structure of the active alleles permits access of factors to interact with their DNA binding sites. It is also possible that hypermethylation of the 5' region of housekeeping genes on the inactive X chromosome may have a role in establishing or stabilizing local chromatin structure of 5' cis-acting regulatory sites (and/or GC islands).

CHAPTER 3
IN VITRO RECONSTITUTION OF A DNA-PROTEIN INTERACTION SPECIFIC TO THE ACTIVE HPRT ALLELE

Introduction

The *in vivo* footprinting of the human HPRT gene on the active and inactive X chromosomes revealed multiple footprints specific to the active X chromosome (41). Of the six footprints specific to the active HPRT allele, four of the footprints occur at GC boxes or potential Sp1 binding sites, one occurs at a potential AP-2 binding site, and the other occurs at a target DNA sequence which appears to represent a newly cis- and trans-acting regulatory element. The footprint of this new DNA-protein interaction consists of a strong DMS reactive sites at position -91 (relative to the translation initiation codon) on the upper strand and at -90 and -75 on the lower strand (termed the -91 footprint). No obvious protections are seen around the -91 footprint.

A DNA data search with the DNA sequence from the immediate region containing the enhanced DMS-reactive sites at position -91 to position -75 did not yield clear sequence identity with any previously described regulatory elements among vertebrate control DNA sequences (24,63). In the human HPRT gene 5' region transcription starts at multiple

sites from -104 to -169 (50,81). The position of the -91 footprint just 3' to the multiple sites of transcription initiation suggests the protein(s) associated with this DNA sequence may function in transcription initiation as has been postulated for other DNA-binding regulatory factors located in a similar position. These factors include HIP-1 (69), Inr (103), YY1 (100), and TFII-1 (97). Comparison of the DNA sequences in the -91 footprint with the DNA sequences bound by these initiation factors yielded no significant sequence similarity between these cis-acting elements and the -91 footprint. This further suggests the DNA-protein interaction(s) in the -91 footprint may represent new regulatory elements involved in transcription initiation.

To characterize the DNA-protein interaction which constitutes the -91 footprint, electrophoretic gel mobility shift assays (25,28) have been performed to reconstitute the DNA-protein interaction *in vitro* using crude HeLa nuclear extracts and cloned DNA fragments containing the -91 footprint. Reconstitution of the -91 footprint DNA-protein interaction may allow the eventual cloning of the protein(s). The *in vitro* reconstitution experiments may define the role of this DNA-protein interaction in the regulation of HPRT gene expression. Furthermore, reconstitution experiments are a necessary prerequisites before *in vitro* characterization of the protein and *in vitro*

transcription assays. These experiments may also provide insight into the transcription initiation of TATA-less genes. Preliminary gel mobility-shift experiments have demonstrated multiple DNA-protein complexes, some of which can be abolished by the addition of excess specific promoter competitors.

Materials and Methods

Nuclear Extracts

Nuclear extracts were prepared from suspension cultures of HeLa S3 cells. HeLa S3 cell were grown in suspension modified minimal essential media with 10 % fetal bovine serum. One to three X 10^9 cells were grown and nuclear extracts were prepared as described by Dignam et al. (17). Crude nuclear extracts were quantified with the Bio-Rad protein assay using bovine gamma globulin as a protein standard.

Preparation of Cloned DNA Fragments for Gel Mobility-Shift Assays

A 103 bp Bsu36I-BssHII fragment of the human HPRT gene containing the -91 footprinted region was prepared as follows (See Figure 3.1). A plasmid, pλ4X8-RB1.8 (100 ug) (81), containing the human HPRT 5' region was digested with

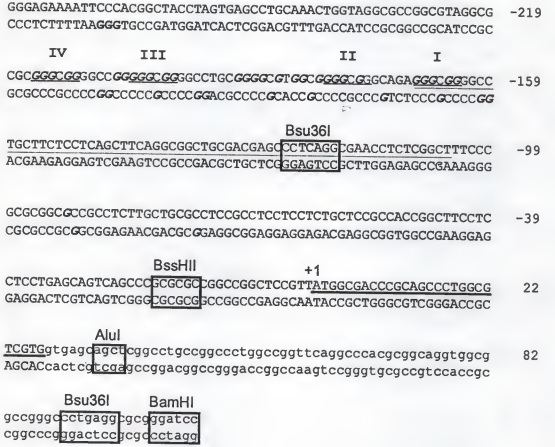

Figure 3.1 - Sequence and Restriction Map of human HPRT 5' region used to prepare cloned DNA fragments for gel mobility-shift assays. The numbers on the right side are relative to the translation initiation codon marked +1. The restriction sites used for the preparation of the subfragments are boxed and indicated above the site. The BamHI site represents the 3' end of the 1.8 kb EcoRI-BamHI fragment cloned into pUC-8 (81). The region of multiple transcription start sites are denoted with the dashed underline. The four GC boxes are thinly underlined and marked I, II, III, IV. Guanine residues footprinted on the active human HPRT gene are shown in bold and italic. The coding region of exon 1 is denoted by the thick underline.

Bsu36I (all restriction enzymes were purchased from New England Biolabs and used according to the manufactures instructions), size fractionated on a 1.6% agarose gel, and the resulting 213 bp Bsu36I fragment isolated from the agarose gel using DEAE cellulose (Schleicher and Schuell) (98). The 213 bp Bsu36I fragment was further digested with AluI and the 157 bp Bsu36I-AluI fragment was separated from the 56 bp AluI-Bsu36I fragments using a 2% agarose gel (Gibco-BRL). The 157 bp Bsu36I-AluI fragment was isolated from the agarose with DEAE cellulose. Next, the 157 bp Bsu36I-AluI fragment was digested with BssHII and the 103 bp Bsu36I-BssHII fragment was isolated from the 54 bp BssHII-AluI fragment using a 2% agarose gel. After size fractionation, the 103 bp Bsu36I-BssHII fragment of the human HPRT promoter was purified from the agarose using DEAE cellulose. After ethanol precipitation, the 103 bp Bsu36I-BssHII fragment was used without further purification.

The following cloned 5' promoter regions were prepared from plasmids for mobility-shift competition assays: a 1.8 kb EcoRI-BamHI fragment of the human HPRT 5' region from plasmid pλ4X8-RB1.8; a 1.4 kb EcoRI fragment of the mouse HPRT 5' region from plasmid pHPT6; a 400 bp FnuDII fragment of the mouse adenine phosphoribosyltransferase (APRT) gene cut from the plasmid with HindIII; a 625 bp SmaI-Sau3A fragment of the mouse dihydrofolate reductase (DHFR) gene cut from plasmid pSS625 with SmaI and HindIII; a 1.7 kb

HincII fragment of the human albumin promoter cut from pUC-18 with SmaI and HindIII; a 1.2 kb SstI fragment of the human factor VIIIC promoter region from plasmid pSP64; a 812 bp EcoRI-BamHI fragment of the human phosphoglycerate kinase (PGK-1) 5' promoter from plasmid pSPT19 (124). The competitor fragments were digested with the appropriate enzymes and separated from the vector by agarose gel electrophoresis. Then, the fragments were purified from the agarose with DEAE cellulose (98). The competitor DNA fragments were quantitated after agarose gel electrophoresis by comparison the fragments ethidium bromide fluorescence which the fluorescence of known DNA standards. The double-stranded Sp1 and AP-2 consensus sequence oligonucleotides were purchased from Promega and a two complementary 18-mers (-83 to -76 of the human HPRT promoter region) were synthesized and annealed using standard techniques (98).

Electrophoretic Gel Mobility Shift Assays

The 103 bp Bsu36I-BssHII fragment was first radiolabelled with ^{32}P-α-dCTP using klenow fragment of DNA polymerase I to fill in the 5' overhang (98). The 20 ul binding reaction (14) consisted of 15000 counts per minute of labelled fragment (0.5 ng), 1 ug [poly (dI:dC)][poly(dI:dC)] as nonspecific competitor, 5 ug of crude HeLa nuclear extract, in 1X binding buffer (12% glycerol, 12 mM HEPES NaOH, pH 7.9, 60 mM KCl, 5 mM MgCl2, 4

mM Tris, pH 8, 0.6 mM EDTA, 0.6 mM dithiothreitol). The binding reaction was allowed to incubate for 20 minutes at room temperature. After incubation, the binding reaction was size fractionated on a 4% acrylamide (80:1 acrylamide:bis-acrylamide) gel containing 50 mM TBE (1 molar TBE = 1 Molar Tris where boric acid is added until the pH is 8.3, 10 mM EDTA). After electrophoresis, the gel was dried and exposed to Kodak XAR film for 1-3 days.

Results

To reconstitute *in vitro* the DNA-protein interaction which comprises the -91 footprint *in vivo* a 103 bp Bsu36I-BssHII fragment of the human HPRT promoter was prepared. This fragment was selected for gel-shift analysis because it contains the -91 footprinted region and some flanking sequence exactly where a specific *in vivo* footprint is seen on the active HPRT allele (41). This restriction fragment does not contain the sequences in the region of the GC boxes which are also footprinted *in vivo* on the active X chromosome. The binding of sequence-specific transcription factors to the cloned DNA fragment containing the human HPRT gene was detected by electrophoretic gel mobility-shift assays (25,28). The cloned fragment was incubated with crude HeLa nuclear extracts (17), resulting DNA-protein complexes were size fractionated on a native acrylamide gel.

Figure 3.2 shows the results of gel mobility-shift

assays using the 103 bp Bsu36I-BssHII fragment of the human HPRT promoter. During the incubation of the cloned DNA fragment with the HeLa nuclear extract, proteins bind to the DNA fragment. After native gel electrophoresis, DNA-protein complexes are visualized as bands with retarded mobility in the autoradiogram. In preliminary experiments, multiple DNA-protein complexes were seen similar to those in Figure 3.2, lane 1. Of the multiple complexes formed, two were of greatest intensity and these are labelled complex I, II in Figure 3.2. Many other complexes were formed but these were of lesser intensity. In initial experiments, the amount of nonspecific competitor (dI-dC) and nuclear extract were optimized for the formation of individual DNA-protein complexes (data not shown). DNA-protein complex formation was shown to increase with increasing amounts of nuclear protein until a threshold where the nonspecific binding of the extract saturates the nonspecific competitor. The amount of nonspecific competitor was optimized to prevent the formation of nonspecific complexes. In Figure 3.2, lane 13 shows the free labelled fragment and lane 1 shows the complexes formed upon the addition of HeLa nuclear extract. In Figure 3.2, lane 2, Multiple bands of retarded mobility are seen, indicating multiple DNA-protein complexes. The pattern of retarded bands is consistent over a wide range of salt conditions (up to 250 mM KCl), and different nonspecific competitors.

Figure 3.2 - Electrophoretic mobility-shift assays using cloned promoter regions fragments from other genes as unlabelled competitor DNA. Lane 1 is the pattern of DNA-protein complexes seen without competitor DNA added. Lane 13 is the free labelled DNA fragment without any protein added. All competitor DNA were added at a 100-fold molar excess except for lane 11 where a 700-fold molar excess was added. Specific competitors were added to the following lanes: lane 2, 1.8 kb fragment of the human HPRT promoter; lane 3, 1.4 kb fragment of the mouse HPRT promoter; lane 4, 812 bp fragment of the human PGK-1 promoter; lane 5, 625 bp fragment of the mouse DHFR promoter; lane 6, 400 bp fragment of the mouse APRT promoter; lane 7, 1.2 kb fragment of the human factor VIIIC promoter; lane 8, 1.7 kb fragment of the human albumin promoter; lane 9, Sp1 consensus oligonucleotide; lane 10, AP-2 consensus oligonucleotide; lane 11, unlabelled 103 bp Bsu36I-BssHII fragment of the human HPRT gene; lane 12, double-stranded 17-mer from position -83 to -76 of the human HPRT promoter.

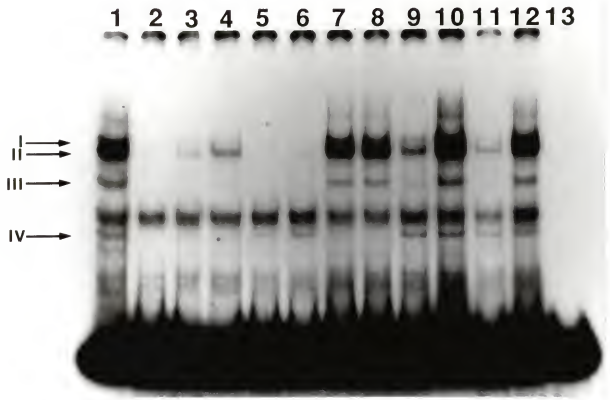

Figure 3.2 - Electrophoretic mobility-shift assays using cloned promoter regions fragments from other genes as unlabelled competitor DNA.

To determine whether or not the retarded bands represent sequence-specific binding of a DNA-binding protein(s), the same mobility-shift assay was performed in the presence of specific competitor DNA fragments. These competitors consisted of 5' promoter regions from housekeeping and tissue-specific genes, double stranded oligonucleotides containing consensus Sp1 or AP-2 binding sites, and a double-stranded oligonucleotide containing a DNA sequence just 3' of the -91 footprint (see materials and methods) were added to the binding reaction in 100-fold molar excess of the labelled fragment.

Results of competition mobility-shift analysis are shown in Figure 3.2, lanes 2-12. In lane 2, a 1.8 kb fragment of the human HPRT promoter region (from which the radiolabelled fragment was prepared) is used as competitor. Addition of the 1.8 kb HPRT promoter fragment abolished complexes I, II, III, and IV. Complexes I and II are the major complexes in the gel mobility-shift assay. Addition of a 1.4 kb fragment of the mouse HPRT promoter region (lane 3) demonstrates similar results to competition with the human promoter except the mouse promoter fragment is less efficient in the abolition of complex II. When the 812 bp fragment containing X-linked human PGK-1 gene was used as a competitor, complexes I, III, and IV are effectively abolished and complex II is greatly reduced (lane 4).

Competitions using fragments from the mouse dihydrofolate reductase (DHFR) and mouse adenine phosphoribosyltransferase (APRT) promoters demonstrate nearly compete competition of all four complexes (lanes 5, 6). Two other promoters fragments, containing the human factor VIIIC and albumin promoter regions, failed to compete significantly complexes I, II, and III but effectively abolished complex IV (lanes 7,8). Addition of a Sp1 consensus double-stranded oligonucleotide to the binding reaction, reduces the intensity of complexes I, II, and III although less efficiently (lane 9). The intensity of complex IV was not altered by the addition of the Sp1 consensus oligonucleotide. However, another GC-rich oligonucleotide containing an AP-2 consensus sequence does not show significant competition of any complex (lane 10). In addition, a double-stranded 17 bp oligonucleotide, containing a DNA sequence just 3' of the -91 footprint, does not significantly compete (lane 12). These data suggest reconstitution of sequence-specific complexes responsible for the in vivo footprint. Alternatively, the data may represent the binding of factors with a specificity toward certain sequences in GC-rich DNA.

When a unlabelled 103 bp Bsu36I-BssHII fragment (which is the same as the radiolabelled fragment) was added to the binding reaction in a 100-fold molar excess minimal competition was seen (data not shown), but when a 700-fold

excess of cold fragment was added, complexes I and III were abolished and complexes II and IV were reduced (lane 11). Thus, it appears the fragment itself is a less efficient competitor of complexes I, II, III, and IV than DNA fragments from housekeeping promoters.

Some retarded complexes appear to be nonspecific (complexes not marked) because they are resistant to competition with all specific competitors. Sequence-specific DNA-protein interactions should be abolished by an excess of specific-competitor that includes the binding site.

Discussion

Preliminary *in vitro* reconstitution experiments, using crude HeLa nuclear extracts and a cloned DNA fragment containing the sequence of the -91 footprint, have demonstrated the formation of multiple DNA-protein complexes in gel mobility-shift assays. Competition experiments have analyzed the specificity of the multiple DNA-protein complexes. Cloned DNA fragments from the human HPRT , mouse HPRT, mouse DHFR, and mouse APRT promoter specifically abolish the formation of complexes I, II, III, and IV. These promoters are all GC-rich housekeeping promoters which lack TATA boxes and are similar in sequence and structure to the human HPRT promoter; this similarity is likely to explain why these fragment are effective competitors. The

mouse HPRT promoter is similar to the human HPRT promoter and contains a 9 bp sequence which exactly matches the -91 footprint. In vivo footprint analysis of the mouse HPRT 5' region has demonstrated a single slightly enhanced DMS-reactive guanine (Litt, Hornstra, and Yang, unpublished data) in the same relative location as -91 footprint in the human HPRT promoter, and this may explain the effective competition of the complexes I, II, III, and IV. The human PGK-1 promoter competes complex II less effectively than the other housekeeping promoters but PGK-1 does not contain sequences matching -91 footprint or in vivo footprints (86) in a similar location as the -91 footprint.

Two tissue-specific promoters, the human factor VIIIC and albumin promoter, do not compete DNA-protein complexes I, II, III, and IV significantly. The lack of competition with two tissue-specific promoters suggests complexes I, II, III, and IV are specific. Initial, competition experiments with a Sp1 consensus oligonucleotide reveals some degree of competition (Figure 3.2, lane 9), although purified Sp1 protein will not bind significantly the Bsu36I-BssHII fragment (data not shown). The Sp1 oligonucleotide may share enough sequence similarity with -91 footprint to compete to a lessor degree. In contrast, a AP-2 consensus oligonucleotide (also GC-rich) does not show any significant competition. Addition of a double-stranded 17-mer which contains human HPRT sequence just flanking the -91

footprint, no competition of any complexes is observed with this fragment. Thus, it appears the site of the DNA-protein interaction is not contained on this small DNA fragment or the fragment contains insufficient flanking sequence for efficient binding. When unlabelled Bsu36I-BssHII fragment is used in competition experiments, complex formation is only slightly inhibited at a 100-fold excess but a 700-fold excess demonstrates significant competition. Thus, the fragment itself competes at low efficiency. The reason for the inefficient competition with the fragment itself is unknown but may be due to the complexity of DNA-protein interaction or the preparation of the nuclear extracts.

These gel mobility-shift assays and competitions experiments are reproducible. The results demonstrate that GC-rich housekeeping promoters compete significantly, but tissue-specific promoters do not complete effectively. The weak competition using the unlabelled Bsu36I-BssHII fragment in mobility-shift assays is puzzling. Current studies are underway to examine subfragments of the human HPRT 1.8 kb promoter region for there ability to act as efficient competitors. Further study of the -91 footprint which has been reconstituted using in vitro DNase I or DMS footprinting may define the exact binding site of this DNA-protein interaction. However, these in vitro footprinting studies may require partial purification of the DNA-binding

protein by heparin-agarose chromatography or affinity chromatography.

CHAPTER 4
HIGH RESOLUTION METHYLATION ANALYSIS OF THE HUMAN HYPOXANTHINE PHOSPHORIBOSYLTRANSFERASE GENE 5' REGION ON THE ACTIVE AND INACTIVE X CHROMOSOMES: CORRELATION WITH GENE SILENCING AND BINDING SITES FOR TRANSCRIPTION FACTORS

Introduction

During early mammalian female embryogenesis, one of the two transcriptionally active X chromosomes is randomly inactivated in the embryo. The inactivation of one X chromosome in each female somatic cell creates a unique system of differential gene expression where a transcriptionally active X chromosome and a transcriptionally inactive X chromosome occupy the same nucleus. The inactivation of genes on one of the two X chromosome in females compensates for the dosage imbalance of X-linked genes between males and females (31,33). The molecular mechanisms that initiate inactivation, propagate the inactivation signal, and maintain this novel system of differential gene expression through subsequent cell divisions are unknown. DNA methylation (47,60,61,74,85,120,126), chromatin structure (48,80,86), DNA-protein interactions (31,68), and DNA replication (30,107) have all been proposed to have roles in this process.

DNA methylation has been widely implicated in the regulation of gene expression in mammalian cells (5,89). In many systems of differential gene expression, hypermethylation of certain sites within or flanking genes, particularly in regulatory regions (4,56), has been correlated with transcriptional silencing (5,89). DNA methylation in mammals occurs at the cytosine residue of CpG dinucleotides to produce 5-methyl cytosine (57). CpG dinucleotides are generally under-represented in mammalian genomes but occur at high frequency within CpG islands. These regions in mammalian DNA carry a high G+C content and are often associated with genes, a feature that has been utilized to identify genes by positional cloning (75,94,95,114). CpG islands are often located in the 5' region of constitutively expressed housekeeping genes and are frequently unmethylated in mammalian DNA (4,5,57). However, CpG islands associated with the 5' region of housekeeping genes on the inactive X chromosome are characteristically hypermethylated (61,86,120,125,126). Numerous studies have examined the role of DNA methylation in the process of X chromosome inactivation. Using a variety of experimental approaches, these studies have investigated a correlation between DNA methylation and maintenance of the transcriptionally silent state of genes on the inactive X chromosome (31,33). These experimental approaches include methylation analysis by methyl-sensitive

restriction enzymes in conjunction with Southern blotting (61,86,110,120,126), DNA-mediated transformation studies using DNA from the active or inactive X chromosomes (60,112), and analysis of the reactivation of genes from the inactive X chromosome using the DNA-demethylating agent 5-azacytidine (37,74,110,113). All support the view that the 5' CpG island of housekeeping genes on the inactive X chromosome are hypermethylated in comparison to their corresponding alleles on the active X chromosome. However, these studies have not established a consistent correlation between specific sites or levels of DNA methylation in the 5' CpG island and transcriptional repression on the inactive X chromosome (47,120,126). Furthermore, a strong correlation between DNA methylation and transcriptional silencing on the inactive X chromosome has not been convincingly established outside of 5' CpG islands (120,126), nor in X-linked tissue-specific promoters (16). The role of DNA methylation in the process of X inactivation appears to be that of stabilizing the transcriptionally inactive state of CpG-rich promoters following the primary inactivation event (62,102).

Despite the strong correlation between DNA methylation and silencing of housekeeping genes on the inactive X chromosome, the mechanism by which DNA methylation may repress gene expression on the X chromosome is unclear. Methylation within cis-acting regulatory elements may

interfere with the binding of trans-activating factors to their target sites on DNA (51,117), but the binding of certain transcription factors such as Sp1 and CTF is unaffected by methylation of their binding sites (3,39,40). Methylated DNA may also be a target for DNA-binding proteins that preferentially interact with methylated DNA, thereby repressing transcription of a methylated promoter (71,72,116). Alternatively, DNA methylation may suppress transcription by altering chromatin structure (13,49). Recent evidence suggests that methylation within the preinitiation domain of the promoter exhibits the strongest correlation with repression of promoter activity (56). Thus, specific sites or regions within the promoter may be crucial for repressing transcription of genes on the inactive X chromosome by DNA methylation.

Recently, Pfeifer et al. (85,86) have examined the methylation of individual cytosine residues in the 5' CpG island of the X-linked human phosphoglycerate kinase (PGK-1) gene. They have employed the high resolution technique of ligation-mediated polymerase chain reaction (LMPCR) genomic sequencing to determine the methylation state of each and every CpG dinucleotide on the active and inactive X chromosome. This method overcomes the significant limitations of methylation analysis using methylation-sensitive restriction enzymes in conjunction with Southern blot analysis. Methylation-sensitive restriction enzymes

assay only a small fraction of all CpG dinucleotides, and often do not permit precise mapping of methylated and unmethylated restriction sites in regions with a high density of closely spaced restriction sites (such as CpG islands), particularly if the region is partially methylated or unmethylated. Genomic sequencing permits direct examination of the methylation state all cytosines regardless of methylation status, and allows determination of a comprehensive high resolution methylation pattern within a specific region of genomic DNA.

To survey the methylation state of each cytosine residue within the 5' CpG island of the human PGK-1 gene, Pfeifer et al. (85,86) performed genomic sequencing using the ligation-mediated polymerase chain reaction (LMPCR). They found the active PGK-1 allele was completely unmethylated at 120 CpG sites on the active X chromosome, but was essentially completely methylated (118 of 120 CpG sites) on the inactive X chromosome. Hypoxanthine phosphoribosyltransferase (HPRT; EC 2.4.2.8) catalyzes the conversion of hypoxanthine and guanine to IMP and GMP, respectively, in the purine salvage pathway. The HPRT gene is constitutively expressed in all cells and tissues throughout development with elevated expression in the central nervous system, particularly, the basal ganglia (104). The HPRT gene is X-linked and transcriptionally silenced on the inactive X chromosome. We have previously

employed in vivo footprinting to identify the positions of multiple sequence-specific DNA-protein interactions specific to the 5' CpG island of the active HPRT allele; no in vivo footprints were detected on the inactive allele (41).

Previous methylation analysis of the human HPRT gene using methylation-sensitive restriction enzymes suggests that, unlike the PGK-1 gene, the 5' CpG island on the inactive X chromosome is not completely methylated (120,126). Therefore, we have analyzed the human HPRT gene 5' CpG island by LMPCR genomic sequencing to determine the methylation state of every cytosine on the active and inactive X chromosomes, and to determine the complete methylation pattern within the CpG island on the active and inactive X chromosomes. This high resolution map of methylated and unmethylated cytosines was then correlated with transcriptional activity of the gene and the pattern of binding sites for transcription factors that interact with the promoter region in vivo (41). We find a nearly complete absence of DNA methylation on active and 5-azaC-reactivated HPRT alleles. The inactive allele is nearly completely methylated at all CpG dinucleotides, except in the region containing four adjacent GC boxes which has been shown by in vivo footprinting to be bound by sequence-specific DNA binding proteins only on the active allele. CpG dinucleotides in this region are either partially methylated or unmethylated in two independent cell lines carrying an

inactive X chromosome. These data provide insight into molecular processes that may be involved in X chromosome inactivation.

Materials and Methods

DNA, Cells, and Cell Lines

DNA samples were prepared from cultures of cell lines previously described (41). Briefly, GM00468 is a normal diploid human male fibroblast cell line containing an active X chromosome. Cell line 4.12 (generously provided by David Ledbetter) is a hamster-human somatic cell hybrid containing only the active human X chromosome in the HPRT-deficient hamster cell line RJK88 (77); RJK88 carries a deletion of the endogenous hamster HPRT gene (27). Cell line 8121 is a hamster-human somatic cell hybrid containing an inactive human X chromosome in a RJK88 hamster cell background (also provided by David Ledbetter). Cell line 8121R9a is a 5-azacytidine (5-azaC) reactivant of 8121 grown from a single hypoxanthine/aminopterin/thymidine (HAT)-resistant colony expressing the 5-azaC-reactivated human HPRT gene. In some experiments, a second 5-azaC reactivant was studied; cell line M22 is a 5-azaC-treated HPRT reactivant of a mouse-human somatic cell hybrid containing an inactive human X chromosome in a murine A9 cell background (generously provided by Barbara Migeon). An additional cell line, X8-6T2, is a hamster-human somatic hybrid cell line containing

an inactive human X chromosome (18,22,36) (generously provided by Stanley Gartler) and grown in D-MEM with 10% fetal bovine serum and 1% penicillin-streptomycin. In some experiments, HeLa S3 cells which contain an active human X chromosome were included.

All somatic cell hybrids containing an active HPRT gene were cultured using standard techniques in Dulbecco's modified Eagle's medium (D-MEM) (Gibco) with 10% fetal bovine serum (FBS), 1% penicillin-streptomycin supplement (P-S; Gibco), and supplemented with 1X HAT (0.1 mM hypoxanthine, 0.4 uM aminopterin, 0.016 mM thymidine). Cultures of cell line 8121 were maintained as above without HAT. Human fibroblasts were maintained in Ham's F-12 (Gibco) with 10-20% FBS and 1% P-S. HeLa cells were grown in suspension using suspension modified essential media (S-MEM) with 5% FBS and 1% P-S.

DNA Preparation and Base-Specific Modification

Genomic DNA from each cell line was isolated as previously described (41). LMPCR genomic sequencing was performed as described by Hornstra and Yang (41). This is a modification of the original genomic sequencing method described by Church and Gilbert (67). Briefly, purified genomic DNA (50 ug) was digested with EcoRI to decrease viscosity, phenol:chloroform (50:50) extracted, and ethanol precipitated. The digested DNA was resuspended in 5 ul

water and 15 ul 5 M NaCl, then subjected to the standard Maxam and Gilbert cytosine-specific modification reaction with hydrazine (67). Hydrazine modification of 50 ug of genomic DNA for 16 minutes at room temperature was found to be optimal. After cleavage of the DNA at hydrazine-modified cytosines by piperidine treatment (67), 1/10 volume of 3 M sodium acetate (pH-7) was added, the DNA precipitated with 2 volumes of ethanol, and collected by centrifugation at 14000 x g for 30 minutes. After decanting the supernatant, the pellet was washed twice with 80% ethanol, and dried overnight in a vacuum concentrator. The chemically cleaved genomic DNA was resuspended in 1 X TE (10 mM Tris pH 8, 1 mM EDTA) at approximately 1 ug/ul.

For controls, 10 ug of plasmid DNA, which contains a 1.8 kb fragment of human HPRT 5' region, was linearized with EcoRI and subjected to each of the four standard Maxam and Gilbert sequencing reactions (G, A+G, T+C, C) (67). After vacuum drying, the plasmid samples were diluted to a final concentration that would produce signals in the final autoradiogram equal in intensity to that of a single copy mammalian gene after LMPCR of genomic DNA.

Ligation-Mediated PCR

LMPCR was carried out as described by Hornstra and Yang (41) with a modification of the Garrity and Wold procedure (29) employing Vent DNA polymerase (New England Biolabs).

For the LMPCR, six primer sets previously described for in vivo footprinting of the human HPRT gene were used (41), as well as two new primer sets, I and J: primer I1 (5'-HO-TTGCTGCGCCTCCGCCTC-OH-3') and primer I2 (5'-HO-CGGCTTCCTCCTCCTGAGCAGTCA-OH-3'); primer J1 (5'-HO-CGCCATTTCCACCTTCTCTT-OH-3') and primer J2 (5'-HO-TTCCCACACGCAGTCCTCTTTTCCCA-OH-3').

For primer extension (first strand synthesis) with Vent DNA polymerase, 1-5 ug of hydrazine- and piperidine-treated genomic DNA (or the equivalent copy number of treated plasmid DNA), 0.6 pmol of primer 1, 3 ul of 5X Vent buffer (5X Vent buffer = 200 mM NaCl, 50 mM Tris-HCl, pH 8.9) were mixed, and water added to bring the total volume to 15 ul. This mixture was incubated at 98°C for 10 minutes to denature the DNA, followed by annealing of the primer at 45°C for 30 minutes. The samples were cooled on ice, and 15 ul of a freshly prepared solution was added to each tube to yield a solution with a final concentration of 40 mM NaCl, 10 mM Tris-HCl, pH-8.9, 5 mM $MgSO_4$, 0.25 mM 7-deaza-dGTP dNTP mix (0.25 mM dATP, 0.25 mM dCTP, 0.25 mM dTTP, 0.1875 mM 7-deaza-dGTP, 0.0625 mM dGTP), and 2 units of Vent DNA polymerase. The first strand synthesis (primer extension) was incubated at 53°C for 1 min, 55°C for 1 min, 57°C for 1 min, 60°C for 1 min, 64°C for 1 min, 68°C for 1 min, 72°C for 3 min, 76°C for 3 min, and then the tubes were placed on ice. Twenty microliters of dilution solution (29) was

added, followed by 25 ul of the ligation solution described by Garrity and Wold (29). The samples were incubated at 17°C overnight for ligation. After the ligation, 40 ul of 7.5 M ammonium acetate and 1 ul of a 10 mg/ml tRNA solution was added to each tube and ethanol precipitated by the addition of 2 volumes of ethanol. The DNA was collected by centrifugation, the supernatant was decanted, the pellet was washed with 80% ethanol, and the pellet was dried under vacuum. The dried pellet was redissolved in 20 ul of water. For PCR amplification, 80 ul of a PCR solution was added so the final concentration in the 100 ul PCR reaction was: 1X Vent buffer, 3 mM $MgSO_4$, 0.25 mM 7-deaza-dGTP dNTP mix, 25 pmole of primer 2, 20 pmole of the 25-mer linker primer, and 3 units of Vent DNA polymerase. Eighty microliters of mineral oil was added to each tube and the samples placed in a temperature cycler (Coy II) for the PCR reaction. The samples were initially denatured at 95°C for 3 minutes, then the tubes repetitively denatured at 95°C for 1 minute, annealed at 66°C for 2 minutes, and extended at 76°C for 3 minutes; the samples were cycled in this manner 20 times. Additionally, with each five cycles, the extension time was increased 30 seconds. After 20 cycles, the tubes were incubated at 76°C and 5 ul of a booster solution (containing 1X Vent buffer, 3 mM MgSO4, 5 mM dATP, 5 mM dCTP, 5 mM dGTP, 5 mM dTTP, and 1 unit of Vent DNA polymerase) was added to each sample. The samples were incubated at 76°C for 10

minutes to allow Vent DNA polymerase to complete the formation of blunt ends. The samples were placed on ice, and 3 ul of 0.5 M EDTA was added. Subsequent gel electrophoresis and electroblotting were carried out as previously described, using a 5% Long Ranger gel (AT Biochem) substituted for the standard polyacrylamide DNA sequencing gel (41). To visualize the final DNA sequencing ladder, single-stranded hybridization probes were synthesized from M13 clones containing the human HPRT promoter region cloned in either orientation. Probe synthesis, hybridization, washing, and autoradiography were performed as previously described (41).

Results

The methylation state of every detectable cytosine in the 5' CpG island of the human HPRT gene was directly examined by genomic sequencing. The 730 bp region spanning positions -530 to +202 (relative to the translation initiation codon) on both the active and inactive X chromosomes was subjected to genomic sequencing analysis using the LMPCR technique (29). This region contains the 5' flanking region, as well as the first exon and the 5' portion of the first intron, and includes most of the 5' CpG island.

The analysis was performed on six different cell lines to examine the methylation state of each cytosine residue on

either the active or inactive HPRT allele. Hybrid cell line 4.12 (77) contains only the active human X chromosome in a hamster cell line that carries a deletion of the HPRT gene (27). Thus, genomic sequencing of DNA from this cell line will determine the state of cytosine methylation on an active human HPRT allele. The active HPRT allele in a diploid human male fibroblast cell line (GM00468) was also analyzed. Cell lines 8121 and X8-6T2 are hamster-human somatic cell hybrids that contain an inactive human X chromosome in HPRT-deficient hamster cell backgrounds (18,22,27,36). Thus, two independently-derived somatic cell hybrids containing an inactive human X chromosome were examined. In addition to the methylation pattern on the active and inactive X chromosomes, the methylation pattern of a 5-azaC-reactivated HPRT gene on the inactive X chromosome was examined in cell line 8121R9a (41). In some experiments, a second 5-azaC-treated HPRT reactivant, M22 (in a mouse A9 cell background), was analyzed. Initially, HeLa cells which contain an active human X chromosome were analyzed but the data is not shown.

Methylation analysis by genomic sequencing (15) is based upon the specificity of the cytosine DNA sequencing reaction of Maxam and Gilbert (67). Hydrazine specifically modifies cytosine residues of genomic DNA in the presence of a high concentration of sodium chloride. Following piperidine cleavage of the DNA at hydrazine-modified

cytosines, the nested set of DNA fragments produced is subjected to electrophoresis on a DNA sequencing gel to generate a cytosine sequencing ladder. However, 5-methylcytosine residues in genomic DNA are resistant to hydrazine modification in the cytosine-specific Maxam and Gilbert reaction. Therefore, 5-methylcytosine residues within genomic DNA appear as missing bands or gaps in the cytosine sequencing ladder when compared to the ladder from an unmethylated sample.

Until recently, it has not been practical to analyze single copy genes in mammalian DNA directly by genomic sequencing because of the high complexity of mammalian genomes. The application of the ligation-mediated polymerase chain reaction (LMPCR) to the original genomic sequencing method of Church and Gilbert (15) now allows direct analysis of purified mammalian DNA (76,85). LMPCR amplifies each DNA fragment in the sequencing ladders from a specific region of interest within genomic DNA after chemical cleavage by the base-specific Maxam and Gilbert reactions. This readily permits direct visualization of the methylation pattern of all cytosines in a specific region of a given gene. The complete set of Maxam and Gilbert DNA sequencing reactions can also be subjected to LMPCR genomic sequencing (from appropriate genomic DNA samples or from plasmid DNA containing the gene of interest) to visualize the complete sequence context of the methylated cytosines.

We have employed this method to examine methylation of the human HPRT gene 5' CpG island on active and inactive X chromosomes.

Methylated cytosines are identified in genomic sequencing autoradiograms by the absence of a band in the cytosine-specific DNA sequencing ladder. For our analysis, an individual cytosine residue was considered to be methylated if the intensity of the band in the sequencing ladder was visually estimated to be less than 25% the intensity of the same band in an unmethylated sample (active X genomic DNA or plasmid DNA containing the human HPRT gene 5' region). Partially methylated cytosines were those that exhibited approximately 25-80% of the unmethylated band intensity, and unmethylated cytosines were those deemed to possess greater than 80% of the control band intensity by visual inspection. Partially methylated sites occur at specific CpG dinucleotides that are methylated in some cells and unmethylated in others within the same cell culture sample.

Figure 4.1 shows the relative positions of the oligonucleotide primer sets and the region covered by each primer set for LMPCR genomic sequencing of the human HPRT gene 5' region. The region between positions -530 to +202 was analyzed for cytosine methylation on both strands. Primer sets N, A, M, and I were used to analyze the lower strand of the HPRT 5' region, and primers sets J, E, C, and

R were used to examine methylation of the upper strand. Cytosine-specific genomic sequencing ladders using primer sets N, A, M, I, J, E, C, and R are shown in Figures 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, and 4.9, respectively.

Analysis of the Lower Strand

Methylation analysis of the 4 CpG dinucleotides between positions -411 to -446 with primer set N yields one unusual methylation pattern (Fig. 4.2). Though all four CpG dinucleotides in the male fibroblast cell line (GM00468) are completely unmethylated (Fig. 4.2, lane 1), 2 of the 4 sites in the hybrid cell line (Fig. 4.2, lane 2) carrying an active X chromosome (4.12) are partially methylated (at positions -425 and -427), and the remaining two sites (positions -411 and -446) are unmethylated. Both cell lines carrying a 5-azaC reactivated HPRT gene (8121R9a and M22) show no methylation at any of these sites (Fig. 4.2, lanes 4,5). On the inactive X chromosome in hybrid 8121, these four CpG dinucleotides are either partially or completely methylated (Fig. 4.2, lane 3); hybrid cell line X8 was not examined in this region. In addition, the active X chromosome in HeLa cells was completely unmethylated in this region (Fig. 4.2, lane 6).

Results of LMPCR genomic sequencing of the lower strand from position -411 to -253 using primer set A is shown in Figure 4.3.

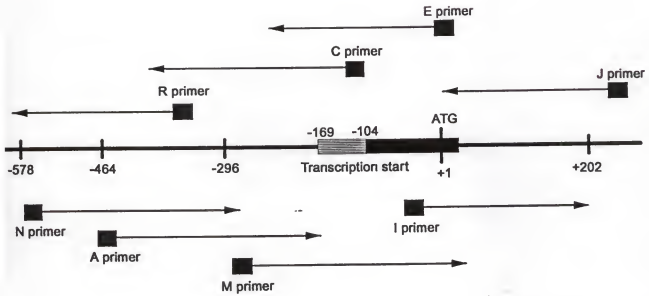

Figure 4.1 - Location of primers used in LMPCR genomic sequencing analysis of the human HPRT 5' region. The numbered line represents the human HPRT 5' region with positions numbered relative to the translation initiation codon. The large rectangle represents the first exon, with the crosshatched portion signifying the region of multiple transcription start sites. The small solid rectangles above and below the numbered line indicate positions of the PCR primers sets used for the LMPCR genomic sequencing. Primer sets N, A, M, and I are complementary to the lower strand sequence and analyze the lower strand; primers E, C, R, and J are complementary to the upper strand sequence and analyze the upper strand. Lines with arrowheads indicate the region resolved by each primer set.

Figure 4.2 - Genomic Sequencing and Methylation Analysis of the Human HPRT 5' Region on the Lower Strand using Primer Set N. The autoradiogram shows the cytosine-specific sequencing ladder from -446 to -411. The position relative to the translation initiation codon is shown to the right of the sequencing ladder. The horizontal bars to the left of the sequencing ladder indicate the position of cytosines in CpG dinucleotides. Genomic DNA from the following sources was used for the genomic sequencing: lane 1, normal diploid male fibroblasts (cell line GM00468); lane 2, hamster-human somatic cell hybrid cells containing the active human X chromosome (cell line 4.12); lane 3, hamster-human somatic cell hybrid cell containing the inactive X chromosome (cell line 8121); lane 4, hamster-human somatic cell hybrid cells containing a 5-azaC-reactivated human HPRT gene on the inactive X chromosome (cell line 8121R9a); lane 5, mouse-human somatic cell hybrid cells containing a 5-azaC-reactivated human HPRT gene on the inactive X chromosome (cell line M22); lane 6, HeLa cells which contain at least an active X chromosome.

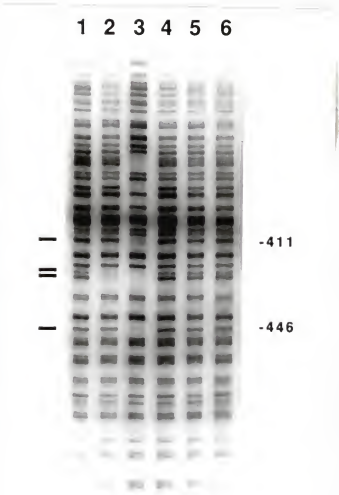

Figure 4.2 - Genomic sequencing and methylation analysis of the human HPRT 5' region on the lower strand using primer set N.

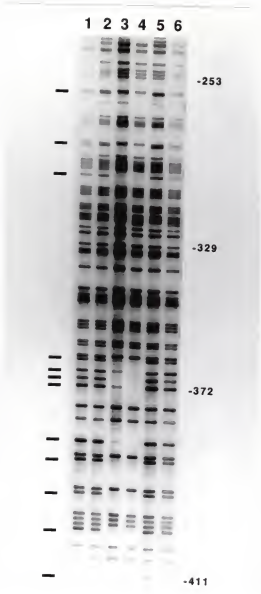

Figure 4.3 - Genomic sequencing and methylation analysis of the human HPRT 5' region on the lower strand using primer set A. The autoradiogram shows the cytosine-specific sequencing ladder from -411 to -253. The symbols and designations are identical to those in Figure 4.2. Genomic DNA from the following sources was used for the genomic sequencing: lane 1, cell line GM00468; lane 2, cell line 4.12; lane 3, cell line 8121; lane 4, hamster-human somatic cell hybrid cells containing an inactive human X chromosome (cell line X8-6T2); lane 5, cell line 8121R9a; and lane 6, cell line M22.

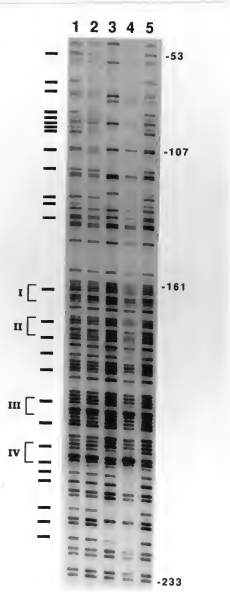

Figure 4.4 - Genomic sequencing and methylation analysis of the human HPRT 5' region on the lower strand using primer set M. The autoradiogram shows the cytosine-specific sequencing ladder from -232 to -53. The symbols and designations are identical to those in Figure 4.2. Genomic DNA from the following sources was used for the genomic sequencing: lane 1, cell line GM00468; lane 2, cell line 4.12; lane 3, cell line 8121; lane 4, cell line X8-6T2; and lane 5, cell line 8121R9a. Methylation data from cell line M22 is not shown. The brackets and roman numerials on the left indicate GC boxes I, II, III, IV.

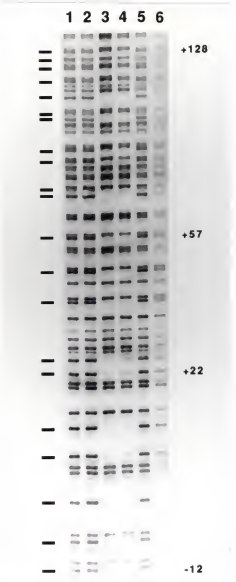

Figure 4.5 - Genomic sequencing and methylation analysis of the human HPRT 5' region on the lower strand using primer set I. The autoradiogram shows the cytosine-specific sequencing ladder from -12 to +128. The symbols and designations are identical to those in Figure 4.2. Genomic DNA from the following sources was used for the genomic sequencing: lane 1, cell line GM00468; lane 2, cell line 4.12; lane 3, cell line 8121; lane 4, cell line X8-6T2; lane 5, cell line 8121R9a; and lane 6, cell line M22.

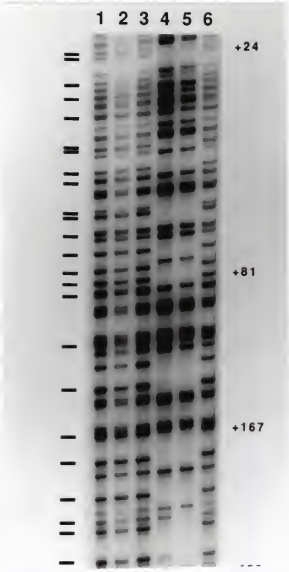

Figure 4.6 - Genomic sequencing and methylation analysis of the human HPRT 5' region on the upper strand using primer set J. The autoradiogram shows the cytosine-specific sequencing ladder from +188 to +24. The symbols and designations are identical to those in Figure 4.2. Genomic DNA from the following sources was used for the genomic sequencing: lane 1, normal male leukocytes; lane 2, cell line GM00468; lane 3, cell line 4.12; lane 4, cell line 8121; lane 5, cell line X8-6T2; and lane 6, cell line 8121R9a.

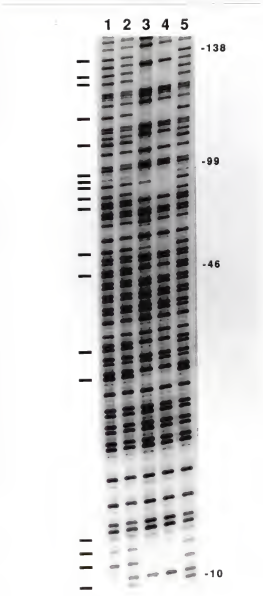

Figure 4.7 - Genomic sequencing and methylation analysis of the human HPRT 5' region on the upper strand using primer set E. The autoradiogram shows the cytosine-specific sequencing ladder from -10 to -134. The symbols and designations are identical to those in Figure 4.2. Genomic DNA from the following sources was used for the genomic sequencing: lane 1, cell line GM00468; lane 2, cell line 4.12; lane 3, cell line 8121; lane 4, cell line X8-6T2; and lane 5, cell line 8121R9a.

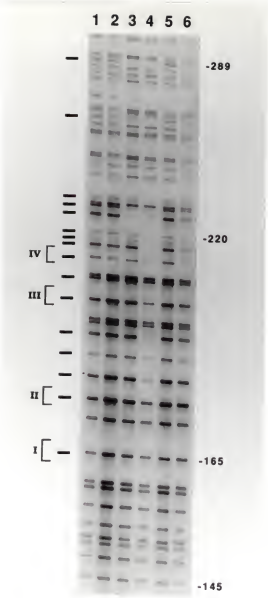

Figure 4.8 - Genomic sequencing and methylation analysis of the human HPRT 5' region on the upper strand using primer set C. The autoradiogram shows the cytosine-specific sequencing ladder from -145 to -289. The symbols and designations are identical to those in Figure 4.2. Genomic DNA from the following sources was used for the genomic sequencing: lane 1, cell line GM00468; lane 2, cell line 4.12; lane 3, cell line 8121; lane 4, cell line X8-6T2; lane 5, cell line 8121R9a; and lane 6, cell line M22. The brackets and roman numerial on the left indicate GC boxes I, II, III, and IV.

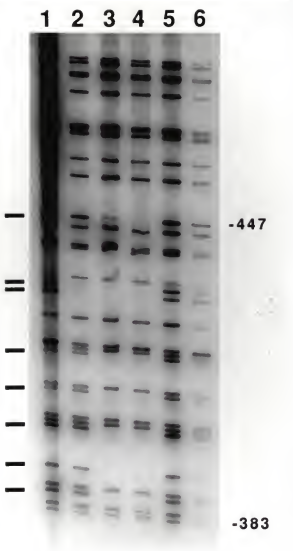

Figure 4.9 - Genomic sequencing and methylation analysis of the human HPRT 5' region on the upper strand using primer set R. The autoradiogram shows the cytosine-specific sequencing ladder from -383 to -447. The symbols and designations are identical to those in Figure 4.2. Genomic DNA from the following sources was used for the genomic sequencing: lane 1, cell line GM00468; lane 2, cell line 4.12; lane 3, cell line 8121; lane 4, cell line X8-6T2; lane 5, cell line 8121R9a; and lane 6, cell line M22.

Within this region, all cytosine residues in CpG dinucleotides are unmethylated in cell lines containing an active HPRT allele (Fig. 4.3, lanes 1 and 2). This was determined by comparing the cytosine band intensities of these samples to those in a similar cytosine-specific LMPCR genomic sequencing ladder from purified plasmid DNA containing the human HPRT 5' region (plasmid ladder not shown); bacterial plasmid DNA is not methylated at CpG dinucleotides. On the active allele, the relative intensity of all cytosine bands from CpG dinucleotides was the same for the plasmid DNA and the genomic DNA samples containing an active X chromosome.

Analysis of the two somatic cell hybrids containing an inactive human X chromosome (Fig. 4.3, lanes 3 and 4) shows hypermethylated cytosines at all CpG dinucleotides in the region covered by this primer set. For example, the cytosine at position -372 displays strong bands in the two samples containing active X chromosomes (lanes 1 and 2) indicating lack of significant methylation, and exhibits significantly less intense bands in the two samples containing an inactive X chromosome (lanes 3 and 4). In cell line 8121 (lane 3), the band intensity is significantly reduced (compared to the unmethylated samples containing the active X in lanes 1 and 2), but is still readily detectable, indicating a partially methylated cytosine at this position in this cell line. However, cell line X8 shows no band

detectable above the faint background ladder (lane 4), indicating that the cytosine at position -372 in this sample is completely methylated. In both cell lines where the inactive human HPRT gene has been reactivated by 5-azaC treatment (8121R9a and M22), the relative band intensity indicative of an unmethylated cytosine is restored (lanes 5 and 6).

Examination of all CpG dinucleotides in this region with primer set A demonstrates that on the active X chromosome (Fig. 4.3, lanes 1 and 2) and in the 5-azaC reactivated HPRT gene (Fig. 4.3, lanes 5 and 6), all cytosines are unmethylated. Analysis of the inactive human X chromosome demonstrates hypermethylation of CpG dinucleotides (primarily fully methylated sites with a few partially methylated sites) in cell line 8121 (Fig. 4.3, lane 3), and complete methylation of all CpG's in cell line X8-6T2 (Fig. 4.3, lane 4).

Results of LMPCR genomic sequencing of the lower strand from position -233 to -53 with primer set M is shown in Figure 4.4. Again, all CpG dinucleotides in this region are unmethylated in both cell lines containing an active HPRT allele (Figure 4.4, lanes 1 and 2), as well as in the 8121R9a 5-azaC reactivant (lane 5) and in 5-azaC reactivant M22 (data not shown). In the two samples containing an inactive human X chromosome, all CpG dinucleotides are completely methylated in the region between positions -53 to

-139 (Fig. 4.4, lanes 3 and 4). However, immediately upstream of this region in these samples, between positions -164 and -233, all CpG's are either completely unmethylated or partially methylated on the inactive X chromosome (lanes 3 and 4). This cluster of hypomethylated sites on the inactive X chromosome coincides with the location of four GC boxes (marked I, II, III, IV in Fig. 4.4) which exhibit in vivo footprints on the active HPRT allele (41). Curiously, no in vivo footprints have been detected in this region on the inactive allele. In cell line 8121, the region containing the four GC boxes (positions -164 to -219) on the lower strand consists entirely of unmethylated sites (Fig. 4.4, lane 3). But further downstream of position -164 in this cell line, nearly all of the CpG dinucleotides return to the completely methylated state. Similarly, in cell line X8-6T2, the same GC box region on the lower strand contains an interspersed pattern of unmethylated, partially methylated, and completely methylated sites (Fig. 4.4, lane 4). Again, further downstream of position -164 in this cell line, nearly all of the CpG dinucleotides are completely methylated.

Results from analysis of the lower strand from position -12 to position +128 using primer set I (Fig. 4.5) indicate both cell lines carrying an active X chromosome (GM00468 and 4.12) as well as both cell lines with a 5-azaC reactivated

HPRT gene (8121R9a and M22) are unmethylated at all CpG dinucleotides. In both cell lines carrying an inactive X chromosome (8121 and X8), methylation of CpG's is nearly complete in this region.

Analysis of the Upper Strand

Analysis of the upper strand from position +202 to +24 was carried out using primer set J Figure 4.6. In this region, the active and 5-azaC-reactivated HPRT alleles again are completely unmethylated at all CpG dinucleotides. The inactive HPRT allele is completely methylated at all CpG's in cell line X8 and methylated at all CpGs in cell line 8121 except at position +186 which is completely unmethylated and position +194 which is partially methylated.

On the upper strand, results of LMPCR genomic sequencing from positions -10 to -138 using primer set E are shown in Figure 4.7. On the active alleles (lanes 1 and 2) and the 5-azaC-reactivated gene in 8121R9a (lane 5), all cytosines in CpG dinucleotides are unmethylated. In both somatic cell hybrids containing an inactive X chromosome, all CpG's are completely methylated (lanes 3 and 4). This region contains an *in vivo* footprint at or near position -91 only on the active allele (41).

The region spanning positions -145 to -289 was examined on the upper strand using primer set C (See Figure 4.8). This region contains the four GC boxes (-164 to -233) which

are denoted on the genomic sequencing ladder in Figure 4.8 as I, II, III, and IV. The GC box region is completely unmethylated on the active (lanes 1 and 2) and 5-azaC-reactivated alleles (lanes 5 and 6). In both cell lines carrying an inactive X chromosome (cell lines 8121 and X8), the pattern of methylation of the GC boxes on the upper strand is similar to that seen on the lower strand. In cell line 8121, most CpG dinucleotides in the GC box region are unmethylated (lane 3), and in cell line X8, the same region shows an interspersion of methylated, partially methylated, and unmethylated sites (lane 4). Upstream of the GC boxes on the upper strand in both of these cell lines, the pattern of hypermethylation typically found on the inactive X chromosome is restored.

Analysis of the upper strand from position -383 to -447 was performed using primer set R as shown in Figure 4.9. All eight CpG dinucleotides in the normal male cell line (GM00468) are unmethylated in this region. However, in the somatic cell hybrid carrying an active X chromosome (4.12), two of the eight CpG's are partially methylated at positions -426 and -428, while the remaining six sites are unmethylated. The two partially methylated sites in this cell line correlate with the position of the partially methylated sites seen on the lower strand in this cell line (see above) using primer set N. Analysis of the inactive HPRT allele in cell lines 8121 and X8 carrying the inactive

human X chromosome, shows either total or partial methylation at every CpG dinucleotide. This region in both cell lines containing a 5-azaC-reactivated HPRT genes is completely unmethylated.

Summary of Methylation Analysis

The methylation pattern of the human HPRT gene on the active X chromosome was examined in two different cell lines. In the diploid male fibroblast (GM00468), the HPRT gene is completely unmethylated at 142 of 142 CpG dinucleotides assayed; the methylation state of 10 additional sites could not be determined because of technical limitations of the LMPCR genomic sequencing (where some cytosines are not resolvable in the sequencing autoradiogram). Somatic cell hybrid 4.12 containing the active X chromosome is unmethylated at 138 of 142 sites, and partially methylated at a cluster of 4 CpG dinucleotides at the far 5' end of the region analyzed (at positions -426 and -428 on the upper strand, and -425 and -427 on the lower strand).

The inactive HPRT allele was examined in two different somatic cell hybrids containing an inactive human X chromosome. The methylation patterns of the inactive HPRT gene in these two cell lines are summarized in Figures 4.10 and 4.11. In cell line 8121, 107 of 142 CpG dinucleotides are completely methylated, 9 CpG's are partially methylated,

and 26 CpG's are unmethylated. Twenty-four of the 26 unmethylated sites in hybrid 8121 are located in the region of the four GC boxes between -233 and -164. In hybrid cell line X8-6T2 122 of 142 CpG dinucleotides are methylated, 12 of 142 CpG's are partially methylated, and 9 CpG's are completely unmethylated. All 9 completely unmethylated sites and 6 of the 12 partially methylated sites are located in the region of the GC boxes in cell line X8-6T2. Thus, in two independent cell lines carrying an inactive human X chromosome, the region containing the four GC boxes is hypomethylated (with unmethylated and partially methylated sites) relative to the surrounding region of the 5' CpG island.

Reactivation of the HPRT gene on the inactive X chromosome by treatment of cells with 5-azaC demethylates all CpG dinucleotides; cell lines 8121R9a and M22 were completely unmethylated at all 142 of 142 CpG dinucleotides in the 5' region. Thus, 5-azaC reactivation of the human HPRT gene on the inactive X chromosome restored the methylation pattern to a pattern indistinguishable from the active HPRT allele.

Figure 4.10 - Summary of the methylation pattern of cytosines from the human HPRT 5' region on the inactive X chromosome. Methylation pattern on the inactive X chromosome in hybrid cell line 8121. The sequence of the human HPRT 5' region is shown. The numbering on the right side of the sequence indicates the position relative to the translation initiation codon marked as +1. The thick solid line underlines the coding region of exon 1. The thin dashed line indicates the region of multiple transcription initiation sites. The GC boxes (which are footprinted on the active HPRT allele) are indicated by a thin solid lined and marked by roman numerals I, II, III, and IV. Guanine residues that are footprinted by dimethyl sulfate (41) on the active HPRT allele are shown in bold italics. Solid filled circles denote methylated cytosine residues. Partially filled circles indicate partially methylated cytosine residues. Open circles represent unmethylated cytosine residues. Question marks indicate cytosine residues which could not be resolved in the sequencing ladder or whose methylation status could not be determined.

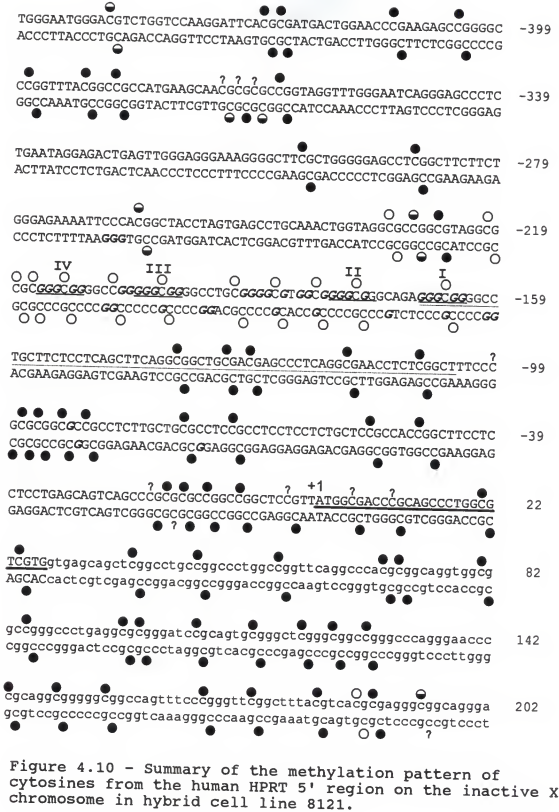

Figure 4.10 - Summary of the methylation pattern of cytosines from the human HPRT 5' region on the inactive X chromosome in hybrid cell line 8121.

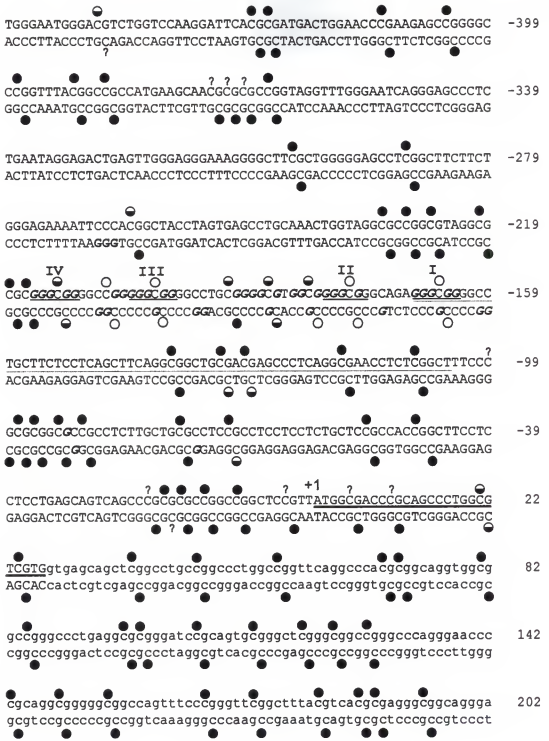

Figure 4.11 - Summary of the methylation pattern of cytosines from the human HPRT 5' region on the inactive X chromosome in hybrid cell line X8-6T2. All symbols are identical to those in Figure 4.10.

Discussion

Methylation analysis of the human HPRT gene by genomic sequencing provides high resolution data that further refines previous methylation analysis by methyl-sensitive restriction enzymes (120,126). Our genomic sequencing studies have focused exclusively on the methylation status of the 5' CpG island and permit an examination of the methylation state of every cytosine nucleotide in the region on active and inactive X chromosomes. This method yields precise and definitive information on the methylation patterns of the active and inactive alleles not available by studies with methylation-sensitive restriction enzymes.

Overall, results from our methylation analysis by genomic sequencing are consistent with previous methylation analysis using restriction enzymes in conjunction with Southern blotting (120,126). These previous studies have indicated that active HPRT alleles are extensively hypomethylated at restriction sites in the 5' region containing the CpG island relative to inactive alleles. However, due in part to technical limitations of these earlier studies, no consistent pattern of methylation at these sites could be discerned and correlated with silencing of the HPRT gene on the inactive X chromosome, particularly within the 5' CpG island. Our analysis by genomic sequencing demonstrates a near total absence of methylation on the active HPRT 5' CpG island in male fibroblast DNA as

well as in a somatic cell hybrid bearing the active human X chromosome and in 5-azaC-reactivated HPRT alleles. The inactive allele in two independent somatic cell hybrids shows a very clear pattern of hypermethylated CpG dinucleotides surrounding a short (48-68 bp) tract of variably hypomethylated sites within the CpG island. These data suggest that some of the heterogeneity in the methylation pattern of the 5' region on inactive HPRT alleles found by using restriction enzymes (120,126) may be due, in part, to this variably hypomethylated region. To date, we have not analyzed the methylation pattern of the 5' CpG island in diploid female cells because of the inability to separate the active and inactive HPRT alleles in these samples.

Correlation of Cytosine Methylation and the Binding of Transcription Factors

In vivo footprint analysis of the human HPRT gene 5' CpG island on the active and inactive X chromosomes has demonstrated multiple footprints specific to the active HPRT allele; no in vivo footprints were detected on the inactive allele (41). The in vivo footprint pattern on the active allele includes evidence for binding of transcription factors to four adjacent GC boxes (positions -163 to -215), DNA sequences shown to interact with the transcription factor Sp1 (7). In addition, the active allele exhibits in vivo footprints at a potential AP-2 binding site (from -265

to -286), and at a position just downstream of the multiple transcription initiation sites that may define the binding site of a new transcription initiation factor (from positions -75 to -91). The positions of these *in vivo* footprints are indicated on Figures 4.10 and 4.11.

On the active HPRT allele, the 5' CpG island is completely unmethylated at all CpG dinucleotides within the DNA sequences of all *in vivo* footprints and in the region of the multiple transcription start sites. This near total absence of methylated cytosines correlates with the binding of transcription factors and transcriptional activity.

The 5' CpG island of inactive HPRT allele, which lacks any evidence for *in vivo* footprints, is extensively methylated. Figures 4.10 and 4.11 present a summary of our methylation analysis of the inactive allele in two different somatic hybrid cell lines carrying an inactive X chromosomes. Comparison of the methylation pattern on the inactive alleles with the pattern of *in vivo* footprints on the active allele reveals an interesting correlation. The region of the 5' CpG island bearing the four adjacent GC boxes is hypomethylated relative to the surrounding regions of the CpG island on the inactive allele, with hybrid cell line 8121 methylated to a lesser extent in this region than hybrid cell line X8-6T2. In cell line 8121, the GC box region is completely unmethylated at all CpG's, while in cell line X8-6T2, the GC box region is interspersed with

unmethylated, partially methylated, and fully methylated CpG's. The molecular basis or cause for this stretch of hypomethylated CpG's on the inactive allele is not clear, though some speculation is possible (see below). This methylation pattern on the inactive allele is particularly unusual because the inactive allele is not bound *in vivo* by sequence-specific binding proteins (41), and the binding of the transcription factor Sp1 to GC box sequences has been shown to be unaffected by CpG methylation within the binding sequence (39,40). Thus, the only hypomethylated region in the 5' CpG island on the inactive allele occurs within unoccupied binding sites for a transcription factor that is not affected by methylation of its DNA target.

One explanation for the hypomethylated GC box region on the inactive HPRT gene may lie in the fact that the region of the four GC boxes has a high incidence of GCG and CGC trinucleotides. DNA methyltransferase may have a bias against methylation of GCG and CGC trinucleotides (86) which would leave these sites hypomethylated in genomic DNA from the inactive HPRT allele. A methylation pattern consistent with this possibility has been noted in the inactive human PGK-1 gene; Pfeifer et al. observed that CGC and GCG trinucleotides are often partially methylated on the inactive PGK-1 allele (86). Examination of the hypomethylated GC box region of the HPRT gene on the inactive X chromosome indicates that unmethylated or

partially methylated sites in this region are often within CGC or GCG trinucleotides. However, these trinucleotide sequences are also frequently found at fully methylated sites on the inactive alleles, and not all unmethylated or partially methylated sites occur within CGC or GCG trinucleotides.

DNA methylation in the GC box region is unlikely to be directly responsible for modulating the differential binding of Sp1 on the active and inactive alleles because this region is hypomethylated on inactive HPRT alleles and because of the ability of Sp1 to bind methylated binding sites (39,40). However, it is possible that methylation could directly affect the binding of other transcriptional activators (at other in vivo footprinted sites on the active allele) by lowering the affinity of the proteins for their binding site on the inactive allele. For example, the in vivo footprinted region involving position -91 (41) is associated with a high density of CpG dinucleotides that are differentially methylated on the active and inactive X chromosomes on both the upper and lower strands; all CpG's in this region are completely unmethylated on the active allele and completely methylated on the inactive allele (see Figures 4.10 and 4.11). It is possible that sequence-specific DNA-binding proteins interacting in this region may be affected by methylation of their binding sites. Methylation of the -91 in vivo footprint region may aid in

repressing transcription of the HPRT gene on the inactive allele because proteins associated with this region may be involved in formation of the preinitiation complex (41), and Levine et al. (56) report that methylation in the preinitiation domain is most effective in suppressing promoter activity.

Further upstream in the 5' CpG island at the potential AP-2 site (or adenoviral E2aE-CB and E4E2 sites) near position -266, two nearby CpGs are also differentially methylated. The two sites are totally unmethylated on the active allele and either partially methylated or fully methylated on the inactive allele. The effect of methylation at this site and in this region is unknown.

Comparison of Cytosine Methylation Patterns on the Human HPRT and PGK-1 Gene 5' Regions

Comparison of the methylation pattern from the human HPRT gene with the pattern obtained by Pfeifer et al. (86) from the X-linked human PGK-1 gene reveals nearly identical patterns on the active alleles. On the active alleles, both genes are unmethylated at CpG dinucleotides; the PGK-1 gene on the active X chromosome is unmethylated at each of 120 CpG's, and the HPRT gene is unmethylated at each of 142 CpG's in male fibroblasts and in 5-azaC-reactivated HPRT genes, and unmethylated at 138 of 142 CpG's in a somatic cell hybrid carrying an active X chromosome. Thus, transcriptional activity of these X-linked housekeeping

genes correlates with an essentially unmethylated 5' CpG island.

On the inactive allele of the HPRT and PGK-1 genes, the general level of methylation is similar (both are hypermethylated relative to the active alleles), but the pattern of methylation is strikingly different. Comparison of the methylation pattern and in vivo footprint pattern of the PGK-1 gene yields no obvious correlation between unmethylated, methylated, or partially methylated sites and the location of binding sites for sequence-specific DNA-binding protein (86). Furthermore, GC box regions in the PGK-1 gene do not show an unusually high incidence of unmethylated or partially methylated sites. However, examination of the human HPRT gene on the inactive X chromosome shows a clear correlation between the GC boxes (which exhibit in vivo footprints only on the active allele) and a cluster of unmethylated and partially methylated sites. It should be noted that the same X8-6T2 hybrid cell line was used in genomic sequencing studies of both genes on the inactive X chromosome. Thus, the difference in methylation patterns between the PGK-1 and HPRT genes on the inactive X chromosome is not simply due to a difference in the cells studied.

Hypermethylation is correlated with the maintenance of transcriptional repression, but as evidenced by the unmethylated and partially methylated sites on the inactive

allele of the HPRT gene, complete methylation of the 5' CpG island is not required for silencing all housekeeping genes on the inactive X chromosome. Thus, the specific position of methylated CpG dinucleotides, the overall density of methylation, and/or the length of methylated regions in the 5' CpG island may be critical for maintaining the transcriptionally suppressed state of housekeeping genes on the inactive X chromosome.

Implications for X Chromosome Inactivation

Hypomethylation of the GC box region on the inactive X chromosome suggests a sequence of events that may occur on the HPRT 5' CpG island early in female embryogenesis at the time of X chromosome inactivation. Transcriptional silencing of the HPRT gene appears to occur prior to de novo methylation of available CpG dinucleotides in the 5' CpG island (62,102). If transcriptional activator proteins bound to the GC box region (most likely Sp1) are not displaced at the time of inactivation of the HPRT gene and prior to de novo methylation of the 5' CpG island, the continued presence of the bound proteins may protect CpG dinucleotides within the binding site from methylation. The delay in displacing transcription factors in the region of the GC boxes in all or some cells during the X inactivation process would allow CpG dinucleotides covered by the binding proteins to escape methylation, resulting in unmethylated

and partially methylated CpG sites in the region occupied by transcription factors on the active allele. Once this pattern of methylation on the inactive X chromosome is established early in embryogenesis, this pattern would persist into adult cells via the maintenance DNA methylase, and would yield the hypomethylated GC box region seen in the two somatic cell hybrids carrying the inactive X chromosome. Presumably, proteins binding to the GC box region would be released or displaced sometime after DNA methylation, since we observed no footprints in this region of the HPRT gene on the inactive X chromosome in our previous in vivo footprinting studies (41).

This scenario would also imply that simultaneous displacement of all transcriptional activators from X-linked genes undergoing inactivation does not necessarily occur at the time of X inactivation, and that displacement of certain key transcription factors may occur first and may be all that is initially required for inactivation of some X-linked genes. In the case of the human HPRT gene, this key factor(s) may be binding to the region surrounding position -91 as seen by previous in vivo footprinting studies (41). This region shows complete methylation in both cell lines carrying the inactive X chromosome. Levine et al. (56) have reported that the most effective repression of genes by DNA methylation was observed when methylation occurred in the preinitiation domain of the promoter. The

position of the -91 footprint region and the absence of a TATA box in the HPRT gene suggests that this DNA-protein interaction may be involved in formation of the preinitiation complex. Thus, displacement of factors from the -91 region followed by methylation of this region by X chromosome inactivation may be sufficient to inactivate the HPRT gene during female embryogenesis. Since there is no evidence for binding of proteins to the analogous region of the active human PGK-1 gene, displacement of Sp1 from the GC box region may be crucial for inactivation of the PGK-1 gene; this could account for the difference in methylation patterns of the GC boxes in the HPRT and PGK-1 5' CpG islands.

Because there is no obvious and consistent correlation between sites of methylated CpG dinucleotides and binding sites for DNA-binding regulatory proteins, methylation of the 5' CpG island of housekeeping genes may be involved in stabilizing the chromatin structure of 5' CpG islands on the inactive X chromosome. This chromatin structure would then be refractory to the binding of transcriptional activators (such as Sp1 and AP-2) and result in transcriptional silencing of the associated genes. This mechanism is supported by 5-azaC reactivation studies of the human HPRT gene by Sasaki et al. (99). These studies indicate that following hemi-demethylation of the HPRT locus on the inactive X chromosome by 5-azaC treatment, a change in

chromatin structure of the HPRT gene precedes reactivation and expression of the HPRT gene. This suggests that DNA methylation may have a role in forming or stabilizing transcriptionally repressed chromatin.

Alternatively, crucial functional sites for DNA methylation may be outside of the 5' CpG island and gene and in a region not analyzed by existing studies. However, the ability to reactivate individual X-linked loci by 5-azaC treatment suggests that, although X chromosome inactivation is a chromosomal process, there is likely to be a some component of regulation at individual loci.

CHAPTER 5
HIGH RESOLUTION METHYLATION ANALYSIS OF THE FMR1 GENE TRINUCLEOTIDE REPEAT REGION IN FRAGILE X SYNDROME

Introduction

The fragile X syndrome is the most common form of inherited mental retardation in man (78). The disease is inherited as an X-linked dominant trait with reduced penetrance and is associated with a folate-sensitive fragile site at Xq27.3. Transmission of the disease within affected families exhibits an unusual pattern of inheritance that includes the existence of transmitting males (101). These males are carriers of the mutation who do not show the disease phenotype. However, grandsons of these transmitting males carry a high risk for expressing the full clinical phenotype of the disease. Abnormal imprinting of the fragile X chromosome by X chromosome inactivation during female embryogenesis has been postulated to be associated with clinical expression of the fragile X mutation (54).

The recent cloning of the FMR1 gene located at the fragile site on the human X chromosome (52,114) indicates that the fragile X syndrome, and the risk of transmitting the disease phenotype, is correlated with the size of a $[CGG]_n$ trinucleotide tandem repeat in the 5' untranslated

region (26). Normal individuals carry allele sizes between 6 and approximately 50 repeat units that are stable upon transmission. Within fragile X families, two classes of increased and unstable repeat numbers are observed. Transmitting males and most unaffected carrier females carry a premutation with a repeat number between 50 to approximately 230. Clinically affected individuals exhibit a major expansion of the premutation repeat number to a full mutation with over 230 repeats, often exceeding 1000. The risk for expansion of the premutation to a full mutation increases with the size of the premutation repeat number, and expansion to the full mutation occurs exclusively during female transmission.

However, expansion of the repeat number to the full mutation is apparently not sufficient by itself to produce the disease phenotype. Expression of the disease phenotype appears to be the result of transcriptional repression of FMR1 gene expression (87). This transcriptional silencing is correlated with methylation of a BssHII within the 5' CpG island containing the CGG trinucleotide repeat, a site not methylated in normal or transmitting males (2,79,115). Methylation analysis with additional methyl-sensitive restriction enzymes also indicate hypermethylation of the repeat and its flanking regions (38). Recently, prenatal diagnosis of a male fetus with fragile X syndrome indicated that fetal tissues show expansion of the trinucleotide

repeat to a full mutation, are methylated at the BssHII site, and show no detectable FMR1 mRNA, while the chorionic villus also carries the full mutation, but is hypomethylated at the BssHII site, and expresses the FMR1 gene (105). Therefore, aberrant methylation at specific sites within the 5' CpG island of the FMR1 gene in affected individuals appears to be correlated with the absence of FMR1 mRNA (and repression of the FMR1 gene) rather than expansion of the repeat number alone. DNA methylation has been widely implicated in gene silencing, particularly in X chromosome inactivation (89). However, the relationship between full expansion of the repeat and DNA methylation, as well as the mechanism by which DNA methylation modulates transcription, are unknown.

The 5' region of the human FMR1 gene that includes the trinucleotide repeat and its immediate flanking regions constitutes a CpG island, a region of mammalian DNA that is unusually high in G+C content and carries a high frequency of the dinucleotide CpG (5). The cytosine residue within CpG dinucleotides (57) is the site at which methylation occurs in mammalian DNA, producing 5-methyl cytosine. However, CpG islands are usually unmethylated in mammalian DNA and are often associated with the 5' region of constitutively expressed genes (4,5). In contrast, hypermethylation of CpG islands is commonly found in the 5' region of genes on the inactive X chromosome in female

somatic cells and is associated with X chromosome inactivation (36,86,110,119,120,126). These hypermethylated 5' CpG islands appear to be a characteristic of many genes on the inactive mammalian X chromosome. This hypermethylation is associated with the transcriptional repression of genes on the inactive X chromosome and has been postulated to stabilize the transcriptionally silent state (84,86).

We have examined the methylation of individual cytosines within and flanking the human FMR1 gene trinucleotide repeat by genomic sequencing (15). This method permits direct methylation analysis of all cytosine residues at single nucleotide resolution in genomic DNA. Thus, the position of every methylated cytosine can be determined within a specific region of interest. This method overcomes the limitations of methylation analysis by methyl-sensitive restriction enzymes (in conjunction with Southern blotting) which is limited by the sequence specificity of the enzymes and their inability to conclusively determine the methylation state of individual CpG dinucleotides in regions with a high density of potential cleavage sites. Using genomic sequencing, we find that all CpGs examined in the immediate flanking regions and within the trinucleotide repeat are completely unmethylated in normal and transmitting males, and methylated in cultured

cells from affected males and in the normal FMR1 gene on the inactive X chromosome.

Materials and Methods

DNA and Cell Lines

DNA samples were obtained from cultures of EBV-transformed lymphoblasts from a normal male, a transmitting male, and an affected male who is the grandson of the transmitting male. DNA samples from normal males were also obtained from blood leukocytes. Cell line 4.12 (generously provided by David Ledbetter) is a hamster-human somatic hybrid cell line containing an active human X chromosome from a fragile X male patient (different from the affected male above). Cell line X8-6T2 is a hamster-human somatic hybrid cell line containing a normal inactive human X chromosome (18,22,36) and was kindly provided by Stanley Gartler.

DNA Preparation and Base-Specific Modification and Cleavage

Genomic DNA was isolated as previously described (41). Purified genomic DNA (50 ug) was digested with EcoRI (an enzyme that does not cleave in the region of interest) to reduce the viscosity of the genomic DNA solutions, phenol:chloroform (50:50) extracted, and ethanol precipitated. The digested DNA was resuspended in 5 ul water + 15 ul 5 M NaCl and subjected to the standard Maxam

and Gilbert cytosine-specific modification/cleavage reaction (67) with hydrazine and piperidine. Hydrazine treatment of 50 ug of genomic DNA for 16 minutes at room temperature was found to be optimal. Following cleavage of hydrazine-modified cytosines by piperidine treatment, 1/10 volume of 3 M sodium acetate (pH-7) was added, the DNA was precipitated with 2 volumes of ethanol, then collected by centrifugation at 14000 x g for 30 minutes. The resulting pellet was washed twice with 80% ethanol and dried overnight in a vacuum concentrator. The chemically-cleaved genomic DNA was resuspended in 1X TE (10 mM Tris pH 8, 1 mM EDTA) at a final concentration of approximately 1 ug/ul. For control samples, 10 ug of plasmid pE5.2 (114), which contains a 5.2 kb fragment of the FMR1 gene including the CGG repeat region, was linearized with EcoRI and subjected to the standard Maxam and Gilbert sequencing reactions (67). After vacuum drying, the plasmid samples were diluted to a final concentration that would produce final autoradiogram signals equal in intensity to that of single copy genes in mammalian genomic DNA after the ligation-mediated polymerase chain reaction (LMPCR).

Ligation-Mediated PCR

LMPCR was carried out as described by Hornstra and Yang (41) using a modification of the Garrity and Wold procedure (29) that employs Vent DNA polymerase (New England Biolabs).

For analysis of the upper strand, primer U1 (5'-HO-CCTAGAGCCAAGTACCTTGT-OH-3') and primer U2 (5'-HO-CACTTCCACCACCAGCTCCTCCATC-OH-3') were used. For the analysis of the lower strand, primer L1 (5'-HO-TTCAGTGTTTACACCCGCAG-OH-3') and primer L2 (5'-HO-CCTAGTCAGGCGCTCAGCTCCGTTT-OH-3') were used. For primer extension (first strand synthesis) with Vent DNA polymerase, 1-5 ug of cleaved genomic DNA (or the equivalent copy number of cleaved plasmid DNA), 0.6 pmol of primer 1, 3 ul of 5X Vent buffer (5X = 200 mM NaCl, 50 mM Tris-HCl, pH 8.9) were mixed and brought to a final volume of 15 ul with water. This mixture was incubated at 98°C for 10 minutes to denature the DNA, followed by annealing of primer 1 at 45°C for 30 minutes. The tubes were cooled on ice, and 15 ul of a freshly prepared solution was added to each tube to yield a final concentration of: 40 mM NaCl, 10 mM Tris-HCl, pH-8.9, 5 mM $MgSO_4$, 0.25 mM 7-deaza-dGTP/dNTP mix (0.25 mM dATP, 0.25 mM dCTP, 0.25 mM dTTP, 0.1875 mM 7-deaza-dGTP, 0.0625 mM dGTP; Pharmacia), and 2 units of Vent DNA polymerase. The first strand synthesis was incubated at 53°C for 1 min, 55°C for 1 min, 57°C for 1 min, 60°C for 1 min, 64°C for 1 min, 68°C for 1 min, 72°C for 3 min, 76°C for 3 min, and then placed on ice. Twenty microliters of dilution solution (29) was added, followed by 25 ul of ligation solution (29). The tubes were incubated at 17°C overnight for ligation. After ligation, 40 ul of 7.5 M

ammonium acetate and 1 ul of a 10 mg/ml tRNA solution were added to each tube and ethanol precipitated by the addition of 2 volumes of ethanol. The DNA was collected by centrifugation, and the pellet washed with 80% ethanol and dried under vacuum. The dried pellet was then redissolved in 20 ul of water.

For PCR amplification, 80 ul of a PCR solution were added to the redissolved DNA sample so the final concentration in a 100 ul PCR reaction were: 1X Vent buffer, 3 mM $MgSO_4$, 0.25 mM 7-deaza-dGTP/dNTP mix, 25 pmole of primer 2, 20 pmole of the 25-mer of the linker primer, 10% glycerol, 5% formamide, and 3 units of Vent DNA polymerase. Eighty microliters of mineral oil were added to each tube and the samples placed in a temperature cycler (Coy II) for PCR. The samples were initially denatured at 98°C for 3 minutes, then the tubes repetitively denatured at 98°C for 20 seconds, annealed at 58°C for 1.5 minutes, and extended at 76°C for 1.5 minutes. The samples were cycled in this manner 20 times. With each cycle, the extension time was increased 5 seconds. After 20 cycles, the samples were incubated at 76°C and 5 ul of a booster solution (containing 1X Vent buffer, 3 mM $MgSO_4$, 5 mM dATP, 5 mM dCTP, 5 mM dGTP, 5 mM dTTP, 10% glycerol, 5 % formamide, and 1 unit of Vent DNA polymerase) was added to each sample. The samples were incubated at 76°C for 10 minutes to allow Vent DNA polymerase to complete the formation of blunt ends on all of

the amplified products. The samples were then placed on ice, and 3 ul of 0.5 M EDTA was added. Gel electrophoresis and electroblotting of the LMPCR-amplified samples were performed as previously described (41). To visualize the sequencing ladder, single-stranded hybridization probes were synthesized from M13 clones containing the CGG repeat in either orientation. Probe synthesis, hybridization, washing, and autoradiography were carried out as described by Hornstra and Yang (41). The radio-labelled hybridization probes were synthesized as described (41) using single-stranded M13 clones containing the 5' region of the FMR1 gene as templates. Clone a51u0001_odd (D.L.N., unpublished data) was used to synthesize the probe specific to the lower strand, and clone a51u0021 was used as the template for synthesis of the probe to analyze the upper strand.

Results

The region within and immediately surrounding the FMR1 trinucleotide repeat was examined by genomic sequencing (15) to determine the methylation state of cytosine residues at single nucleotide resolution. Genomic DNA from normal males, a transmitting male, an affected male (the grandson of the transmitting male), a human-hamster somatic cell hybrid containing an active human fragile X chromosome, and a rodent-human hybrid cell line containing a normal inactive human X chromosome was isolated and subjected to methylation

analysis by LMPCR (ligation-mediated polymerase chain reaction) genomic sequencing. The DNA samples were first treated with hydrazine and piperidine in a standard Maxam and Gilbert cytosine-specific modification and cleavage reaction (67) to generate a cytosine-specific DNA sequencing ladder. Because 5-methylcytosine (5-meC) is resistant to hydrazine modification relative to the reactivity of cytosine with hydrazine, this differential reactivity permits the identification of cytosine residues within mammalian genomic DNA that are methylated. To detect the hydrazine-resistant 5-meC nucleotides, the hydrazine-modified and piperidine-cleaved genomic DNA fragments from the FMR1 repeat region in each genomic DNA sample were amplified by LMPCR, fractionated on a standard DNA sequencing gel, electrotransferred to a nylon membrane, and visualized by hybridization of the membrane with a radiolabelled FMR1 DNA probe followed by autoradiography (41,76,85,86). Methylated cytosines appear as gaps in the final cytosine-specific sequencing ladder when compared to an identical ladder of unmethylated samples. The unmethylated control sample typically employed was plasmid DNA containing the region of interest since *E. coli* DNA is not methylated at cytosines of CpG dinucleotides.

Figure 5.1 shows a diagram of the region within and immediately surrounding the FMR1 gene trinucleotide repeat. The diagram indicates the positions of the two LMPCR

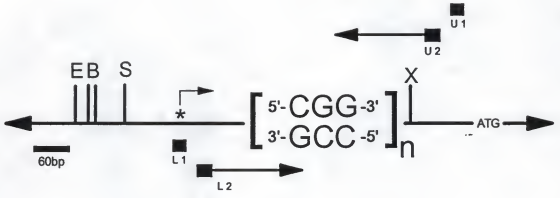

Figure 5.1 - Location of primers used for the genomic sequencing of the human FMR1 gene repeat region. The long horizontal line represents the human FMR1 5' region with the trinucleotide repeat shown in brackets. The asterisk denotes the site of a major transcription start site (S.T.W.; unpublished data) with the bent arrow indicating the direction of transcription. ATG denotes the translation start site. The vertical lines indicate the positions of restriction sites where E = EcoRI, B = BssHII, S = SacII, and X = XhoI. The small solid rectangles above and below the line denote the positions of oligonucleotide primers used in the LMPCR genomic sequencing analysis. Primer set U is complementary to the upper strand, and primer set L is complementary to the lower strand. Arrows extending from the small rectangles indicate the region and direction resolved by each primer set. A 60 bp scale bar is shown below the line.

oligonucleotide primer sets used for this study relative to the position of the trinucleotide repeat region. Each primer set permits examination of one strand of the region within and flanking the repeat. Primer set L anneals to the lower strand and was used to determine the methylation pattern of the lower strand upstream of the trinucleotide repeat and extending into the repeat itself. Primer set U anneals to the upper strand and was used to analyze methylation of the upper strand downstream of the trinucleotide repeat and extending into the repeat. Because of the length of the trinucleotide repeat in some of the samples, it was not possible to examine methylation of the entire repeat. Furthermore, it was not possible to determine the methylation pattern of both strands in the immediate flanking regions because the primers sets required for analysis of the upper strand upstream of the repeat and the lower strand downstream of the repeat would have to anneal to the repeat itself. Primers complementary to the repeat sequence would not anneal to a single position within the FMR1 gene and would not yield specific sequencing ladders after LMPCR.

Analysis of the Lower Strand

Figure 5.2 shows the results from analysis of the lower strand using primer set L. Comparison of the cytosine-specific sequencing ladder--as well as genomic sequencing

Figure 5.2 - Genomic sequencing and methylation analysis of the trinucleotide repeat and immediate flanking region on the lower strand using primer set L. The autoradiogram shows the cytosine-specific sequencing ladder from +88 to +162 in the flanking region, and extending into the repeat region. The positions relative to the transcription start site are shown on the right side of the sequencing ladders (the trinucleotide repeat itself is not included in the numbering). The sequencing ladder proceeds 3' to 5' from the bottom to the top of the figure. The closed circles on the left side of the sequencing ladder represent the position of cytosine in each CpG dinucleotide. The region of the 5'-CCG-3' repeat is indicated by the bracket on the left side of the figure. Genomic DNA from the following sources was used for genomic sequencing: lane 1, normal human male leukocytes; lane 2, transmitting male lymphoblasts; lane 3, affected male lymphoblasts (the grandson of lane 2); lane 4, somatic cell hybrid containing the fragile X chromosome from an affected male (cell line 4.12); lane 5, somatic cell hybrid containing a normal inactive X chromosome (cell line X8-6T2).

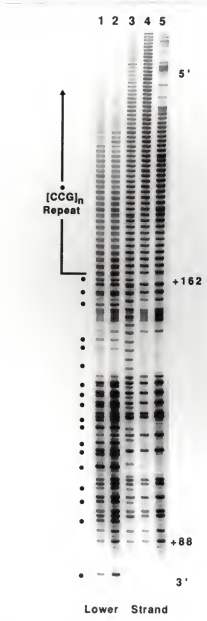

Figure 5.2 - Genomic sequencing and methylation analysis of the trinucleotide repeat and immediate flanking region on the lower strand using primer set L.

ladders from the other Maxam and Gilbert base-specific cleavage reactions (G, G+A, T; data not shown)--with the published nucleotide sequence of this region (26) indicates the sequence corresponding to Figure 5.2 is identical to that of the published sequence, with one exception (Lane 3; see below). The upper portion of the sequencing ladder (within the open bracket) displays the methylation status of the trinucleotide repeat itself. On the lower strand, the sequence of the repeat is $[5'\text{-CCG-}3']_n$, a sequence that contains two cytosines with one CpG dinucleotide in each trinucleotide repeat unit. If the cytosine in the CpG dinucleotide within each repeat unit is not methylated, the repeat unit will appear as a doublet band in the cytosine sequencing ladder with each unmethylated cytosine represented by each of the bands. If the cytosine in the CpG dinucleotide of the repeat unit is methylated, only the first unmethylated cytosine in the repeat unit will be detected in the cytosine-specific sequencing ladder and the repeat unit will be represented as a single band.

As shown in Figure 5.2, in both the normal and transmitting males, the cytosine sequencing ladder within the trinucleotide repeat region displays a continuous ladder of doublet bands, indicating that each and every trinucleotide repeat unit in these samples consists of an unmethylated cytosine doublet. Thus, in both normal and transmitting males, the entire trinucleotide repeat--as far

as can be detected in our autoradiograms--is predominantly or entirely unmethylated. On the other hand, in the affected male, in the affected fragile X human-hamster hybrid, and in the hybrid cell line containing the normal inactive human X chromosome, the cytosine sequencing ladder within the repeat region is a ladder of single bands, indicating that the CpG dinucleotide within every repeat unit is predominantly or entirely methylated. However, within two of the methylated samples (affected male and normal inactive X hybrid; Fig. 5.2, lanes 3, 5) a few sporadic doublets are present within the ladder of single bands. These doublets may indicate occasional repeat units with unmethylated CpG dinucleotides, or more likely, represent the occasional AGG triplet reported to occur within the $[CGG]_n$ trinucleotide repeat (26,114). It is interesting to note that these doublets are very rare, or not observed at all, within the repeat of affected males. For example, in the affected fragile X chromosome hybrid (Fig. 5.2, lane 4) the cytosine ladder can be read clearly enough to determine that no doublet bands are present within the first 80 repeat subunits, and in the affected male (Fig. 5.2, lane 3), only one doublet is detected in the first 80 repeat units. In the normal inactive X hybrid cell line, two doublets are seen within the cytosine-specific ladder of the repeat at a 10 repeat interval and may represent AGG

triplets similar to those seen in the previously sequenced alleles (26,79,114).

In the immediate flanking region of the lower strand shown in Figure 5.2, the cytosine is unmethylated (band in autoradiogram is present) at every CpG examined in normal and transmitting males (lanes 1, 2). In contrast, the same CpGs are completely methylated (band in autoradiogram is missing) in the affected male, the fragile X somatic cell hybrid, and the normal inactive X hybrid (lanes 3, 4, 5; the pattern in lane 3 is complicated by an apparent DNA rearrangement described below). Thus, the complete methylation pattern on the lower strand shown in Figure 5.2 indicates that every CpG dinucleotide in normal and transmitting males is hypo- or unmethylated, while affected fragile X males (in both diploid human cells or in a somatic cell hybrid) as well as the normal FMR1 gene on the inactive X chromosome appear to be completely methylated.

Figure 5.2 also shows a distinct and notable feature of the immediate 5' flanking region upstream of the repeat. A region approximately 18 bases long adjacent to the repeat (from positions +129 to +147) appears very faint in the autoradiogram, a consistently reproducible feature. This region also appears faint after LMPCR genomic sequencing with each of the other Maxam and Gilbert base-specific modification and cleavage reactions (67). These results suggest this region is either relatively resistant to

chemical modification by all of the Maxam and Gilbert modification reagents (dimethyl sulfate, formic acid, and hydrazine), or the 5' end of DNA fragments terminating in this region are less efficiently joined to the linker in the ligation step of the LMPCR procedure. The weak intensity of bands in this region is not due to the failure of the PCR reactions to extend through this region because visualizing the sequence within the trinucleotide repeat (using primer set L) requires that the reaction span this region. This unusual pattern in the autoradiograph may reflect the formation of an novel DNA structure in this region.

In addition, this same region appears to have undergone a rearrangement in a subpopulation of lymphoblast cells from the affected fragile X male (Fig. 5.2, lane 3). This can be seen by the distinct ladder of single bands representing the trinucleotide repeat that extends into the faint region in this sample. The repeat ladder in this patient also appears to continue further into the flanking region upstream (in the 3' direction) of the faint region. However, elements of the normal sequence also appear to be present in the ladder such as the CCCCC sequence around position +148. In addition, further upstream near position +92 in the 3' direction, the normal ladder pattern appears to be restored. Thus, the overall ladder pattern of this region in the affected male appears to consist of two sequencing ladders superimposed upon one another; one is the normal sequence

(as shown by samples in lanes 1, 2, 4, and 5), and the other a rearranged sequence. This suggests that the DNA rearrangement has taken place in a significant subpopulation of cultured lymphoblast cells from this patient. The pattern of the rearrangement is consistent with a small deletion that has occurred immediately flanking, or within the trinucleotide repeat, and extending to a region near position +92. We cannot determine at this time whether or not there is a correlation between the unusual nature of the DNA sequencing ladder in this region and the apparent rearrangement seen in the fragile X sample in lane 3.

Analysis of the Upper Strand

Figure 5.3 shows a similar analysis of the upper strand in the flanking region immediately downstream of the repeat and extending into the repeat. Generating the cytosine-specific ladder by LMPCR genomic sequencing with primer set U, the upper portion of the ladder (within the open bracket) again indicates the methylation status of the trinucleotide repeats. On the upper strand, the sequence of the repeat is $[5'\text{-CGG-}3']_n$, a sequence that contains one CpG dinucleotide in each repeat unit. If the CpG dinucleotide of each repeat unit is not methylated, the repeat unit will appear as a single band in the cytosine sequencing ladder with the unmethylated cytosine represented by the single band. If the cytosine in the CpG dinucleotide of the repeat is

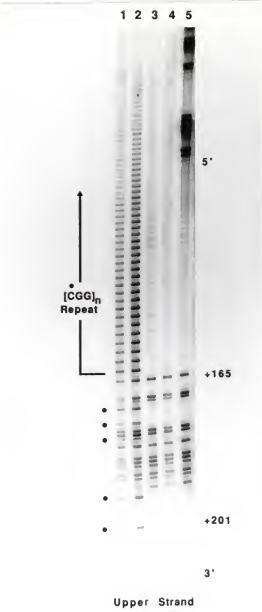

Figure 5.3 - Genomic sequencing and methylation analysis of the trinucleotide repeat and immediate flanking region on the upper strand using primer set U. The autoradiogram shows the cytosine-specific sequencing ladder of the upper strand from positions +201 to +165, and extending into the repeat region. All lane designations and symbols are identical to those in Figure 5.2. However, on the upper strand the repeat sequence is 5'-CGG-3'.

methylated, then no band will appear in the cytosine-specific sequencing ladder (since no other cytosines are present in the repeat unit of the upper strand sequence). As shown in Figure 5.3, in both normal and transmitting males (lanes 1 and 2), the cytosine sequencing ladder within the repeat displays a continuous ladder of single bands corresponding to an unmethylated cytosine within each and every trinucleotide repeat unit. In the affected male, the affected fragile X human-hamster hybrid, and the hybrid cell line containing the normal inactive human X chromosome (lanes 3, 4, 5), only a very faint ladder of bands is detectable within the repeat region, indicating that the CpG dinucleotide within every repeat unit on the upper strand is predominantly or entirely methylated at the cytosine. It is not possible to determine whether the very faint ladder seen in these latter samples (lanes 3, 4, 5) is due to background intrinsic to the LMPCR genomic sequencing technique, or due to very low levels of unmethylated CpG dinucleotides in these samples. The strong bands seen near the top of the lane containing the normal inactive X chromosome (lane 5) represent cytosines on the other side (upstream side) of the repeat.

These results are identical to those found on the lower stand and in the upstream flanking region shown in Figure 5.2. That is, every CpG dinucleotide examined in normal and transmitting males is hypo- or unmethylated, and

hypermethylated or completely methylated in affected males (both in diploid human cells and in a somatic cell hybrid) as well as on the normal inactive X chromosome.

The results from our methylation analysis of both strands are summarized in Figure 5.4. The figure shows the position of each methylated CpG dinucleotide we observed within and flanking the FMR1 repeat of the affected fragile X chromosomes and the normal inactive X chromosome. In these samples (lanes 3, 4, and 5 of Figures 5.2 and 5.3), every CpG that was examined was hypermethylated, whereas no CpGs in normal and transmitting males (lanes 1 and 2 of Figures 5.2 and 5.3) showed detectable methylation.

Discussion

Previous methylation studies of the region surrounding the FMR1 trinucleotide repeat using methylation-sensitive restriction enzymes and Southern blot analysis (2,79,96,106,115) suggested that the FMR1 gene in affected males, but not normal or transmitting males, may be highly methylated. However, due to the limited number of CpG dinucleotides assayed by restriction enzyme analysis, these studies could not determine the complete extent of methylation at all CpGs within and flanking the FMR1 gene trinucleotide repeat. A similar study using methylation-sensitive restriction enzymes that recognize nucleotide sequences within the repeat indicated that the repeat itself

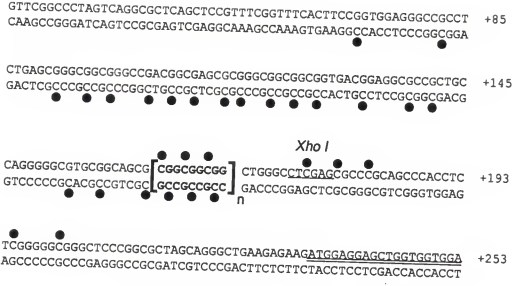

Figure 5.4 - Summary of the methylation state of cytosines from the human FMR1 gene repeat region in affected males and the normal FMR1 gene on the inactive X chromosome. Numbering of nucleotides begins with +1 at the transcription start site indicated in Figure 1. The bracketed region represents the trinucleotide repeat and is not included in the numbering. The double underlined region denotes the protein coding region. An Xho I site downstream of the repeat region is indicated and underlined. Methylated cytosines in the region analyzed are shown as closed circles. The methylation analysis was carried out only on one strand in the regions flanking the trinucleotide repeat (see text for explanation).

was heavily methylated in fragile X patients and in the same X8-6T2 hybrid (containing the normal inactive human X chromosome) used in our studies (38). However, the method used in these studies cannot determine definitively the methylation state of CpGs within each and every restriction site in regions with a high density of closely spaced sites, particularly in samples where sites may be unmethylated or partially methylated.

LMPCR-mediated genomic sequencing (85,86) now permits direct high resolution analysis of the methylation state of individual cytosine nucleotides within and flanking the FMR1 trinucleotide repeat. Using this method, we find the cytosine in all CpG dinucleotides analyzed in this region from affected fragile X chromosomes and from a normal inactive X chromosome to be fully methylated. Cytosine nucleotides from normal males and a transmitting male show very little or no methylation in this region. The extensive methylation of this region of the FMR1 gene in affected patients is very similar to the methylation pattern seen by genomic sequencing of the 5' CpG islands of the X-linked human phosphoglycerate kinase (PGK-1) and human hypoxanthine phosphoribosyltransferase (HPRT) genes on the normal inactive X chromosome. The CpG island in the PGK-1 gene is hypermethylated at 118 of 120 cytosines examined on the inactive X chromosome, whereas the PGK-1 allele on the active X chromosome is completely unmethylated (86).

Genomic sequencing analysis of the human HPRT gene also shows a lack of methylation at all 142 CpG dinucleotides examined in the 5' CpG island on the active X chromosome, and methylation of most (but not all) CpG's in the same region of the inactive allele.

The extensive methylation in the 5' CpG island of genes on the inactive X chromosome is believed to be involved in their transcriptional silencing (84,86,90). Because this pattern of extensive DNA methylation of 5' CpG islands appears to be characteristic of the inactive X chromosome (36,86,110,119,120,126), our methylation analysis of the FMR1 gene suggests that the pattern of hypermethylation seen in fragile X males may be related to X chromosome inactivation. This is supported by the observation of hypermethylation at every CpG dinucleotide examined in the normal FMR1 gene on the inactive human X chromosome. Thus, transcriptional repression of the FMR1 gene in affected fragile X males may involve elements of X chromosome inactivation. Laird has postulated that hypermethylation and silencing of the FMR1 gene in affected fragile X males may be due to aberrant imprinting and failure of the inactive fragile X chromosome to reactivate during gametogenesis (54) in their mothers. However, this would require that the X chromosome carrying the fragile X mutation be selectively inactivated in the female germ line during embryogenesis since 100% of premutations of the FMR1

gene above a threshold of 90 repeat units have been found to expand to the full mutation in oogenesis (26). Alternatively, the methylation patterns observed in the FMR1 gene may suggest that transcriptional repression of the FMR1 gene in fragile X males occurs by a mechanism similar to that used for transcriptional silencing of X-linked genes on the inactive X chromosome, but is not be due directly to the process of X inactivation.

X chromosome inactivation could also contribute to the variable penetrance of the disease in affected females. Random inactivation of either the normal or fragile X chromosome in a crucial subpopulation of cells could result in variable expression of the fragile X phenotype in females carrying the full mutation.

Our data also indicate that the DNA methylation pattern of the mutated FMR1 gene from affected males and from the normal FMR1 gene on the inactive X chromosome is stable in human-hamster somatic cell hybrids. This is demonstrated by the identity of the methylation pattern in these cells to that of cultured lymphoblasts from fragile X males. Hansen et al. (38) have observed partial methylation of certain sites in lymphocytes from fragile X males using analysis with methyl-sensitive restriction enzymes. However, methylation analysis of the human HPRT gene (120,126) indicates that complete methylation of the 5' CpG island is

not required for silencing of the gene on the inactive X chromosome.

Unlike the expanded trinucleotide repeat sequences associated with other genetic diseases such as myotonic dystrophy (9,65), spinal and bulbar muscular atrophy (53), and Huntington's disease (108), the FMR1 repeat associated with the fragile X syndrome is the only repeat which contains a CpG dinucleotide, and it is hypermethylated in affected patients. The mechanism by which expansion and methylation of the trinucleotide repeat affect expression of the FMR1 gene is unknown, though DNA methylation has been shown to stabilize alternative DNA structures such as Z-DNA (1) and triplex DNA (55,66,88). However, the FMR1 repeat sequence does not appear to be capable of forming a Z-DNA structure. DNA methylation has also been postulated to modulate gene expression by affecting the organization of chromatin structure (13,49), as well as affecting the binding of transcription factors to their target DNA sequences (51,117). Further studies of DNA methylation, X chromosome inactivation, alternative DNA structures, and chromatin organization are likely to provide insight into the molecular mechanism of the fragile X syndrome.

CHAPTER 6
CONCLUSIONS AND FUTURE DIRECTIONS

In this dissertation, many aspects of the basic biology of X chromosome inactivation have been investigated. In Chapter 2, the *in vivo* footprint analysis of the human HPRT gene has demonstrated multiple *in vivo* footprints specific to the active HPRT allele while no *in vivo* footprints are observed on the inactive HPRT allele. The *in vivo* footprinting results on the human HPRT gene are similar to the footprinting results of the human X-linked PGK-1 gene (83,86). The footprinting results of these two X-linked genes does not appear to support the hypothesis that X-chromosome inactivation is a process regulated by a specific DNA sequence that binds either activator or repressor proteins within the promoter region of each X-linked gene subject to inactivation (68). The absence of DNA-protein interactions on the inactive allele of the HPRT and PGK-1 genes argues against the presence of a sequence-specific repressor protein which coordinately silences genes on the inactive X chromosome. The data do not support the existence of a sequence-specific activator to potentiate transcription from the active X chromosome since a novel *in vivo* footprinted DNA sequence common to the active alleles

of both genes was not identified. However, one can not exclude the possibility that unique DNA-binding proteins and regulatory sequences specific to X-linked genes subject to inactivation may be located outside of the promoter regions studied.

On the active HPRT allele there are multiple transcription factors bound while on the inactive HPRT allele appears devoid of DNA-binding proteins. Although the active and inactive X chromosomes are located within the same female nucleus, transcription factors are differentially bound. Thus, it appears the inactive X chromosome is inaccessible to the stable binding of transcription factors. The inaccessibility of the inactive X chromosome appears to be related to physical differences when compared to the active X chromosome. These physical differences include DNA methylation on the inactive X chromosome of 5' GC islands of constitutively expressed X-linked genes (36,37,61,74,86,110,119,120,126), and a general decrease of nuclease sensitivity of genes on the inactive X chromosome (36,48,59,122,123). These physical differences have been called differences in chromatin or chromatin structure. In addition to these physical differences, chromatin on the inactive X chromosome is temporally different, being late replicating (30,35).

As a logical extension of the in vivo footprinting studies, Chapter 3 has described preliminary experiments to

reconstitute -91 footprint using gel mobility-shift assays. This in vivo footprint is present in both the human and mouse HPRT genes (Litt, Hornstra, and Yang, unpublished data). The results of the in vitro reconstitution studies using a cloned DNA fragment of the human HPRT gene containing the -91 footprint and crude HeLa nuclear extracts have demonstrated the formation of multiple DNA-protein complexes. The -91 footprint may represent the binding of an initiator element in the TATA-less HPRT promoter. Four of the complexes appear to be specific when competition gel mobility-shift assays are performed. However, these DNA-protein complexes are not efficiently competed by an excess of unlabelled fragment. Further experiments need to be performed to resolved this conflicting data. In vitro footprinting studies may be useful to determine precisely the binding site which would allow the design of specific oligonucleotide substrates. However, if in vitro footprinting results are equivocal then in vivo footprinting of the -91 footprinted region on the active X chromosome using DNase I and LMPCR may allow the confirmation of whether the -91 enhancement represents a DNA-protein interaction or the enhancement is a phenomena secondary to active transcription.

In Chapter 4, methylation analysis of the human HPRT 5' region was performed on the active and inactive X chromosomes using genomic sequencing and LMPCR. The results

demonstrate the absence of methylation at CpG dinucleotides on the active HPRT allele and hypermethylation on the inactive HPRT allele. Curiously, the region of the HPRT promoter that contains the four GC box sequences is hypomethylated on the inactive X chromosome. The mechanism which results in the hypomethylated patch on the inactive allele is unknown. One possible explanation of the hypomethylated patch in the inactive HPRT allele is the sites which are hypomethylated occur at GCG or CGC trinucleotides. These trinucleotides may be poorer substrates for DNA methyltransferase. The region of the GC boxes is bound with transcription factors on the active X chromosome. However, both X chromosome are active during early female embryogenesis before X chromosome inactivation occurs in the late blastocyst. Thus, the hypomethylation of the GC box region may represent the binding of transcription factors when methylation on the inactive X chromosome was established which prevented the methylation machinery from interacting with the GC box region.

Comparison the methylation pattern of the human HPRT 5' region and the human PGK-1 5' region (85,86) demonstrates extensive similarity on the inactive X chromosome. Both genes are hypermethylation in the 5' region on the inactive X chromosome, however, the HPRT 5' region has a patch of hypomethylation not seen in PGK-1. Thus, it appears that DNA methylation of the 5' regions of constitutive X-linked

genes on the inactive X chromosome is important to maintain transcriptional inactivity.

Our data and that of others (83,85,86) advocate the hypothesis chromatin structure and/or DNA methylation may be in part responsible for the differential binding of transcription factors to the active and inactive X chromosomes. DNA methylation of the 5' promoter region on the inactive X alleles could alter the stability of specific-DNA protein interactions to prevent transcription factors from binding. Although, this may be the mechanism modulating the binding of some transcription factors, the data do not support a role DNA methylation in the initiation of X chromosome inactivation (31,33,64). However, in non-eutherian mammals, there is no correlation between hypermethylation and genes on the inactive X chromosome (45). DNA methylation may also alter local chromatin structure and prevent transcription factor access to the inactive allele. Thus, data suggest that chromatin structure and DNA methylation are linked together to possibly inactivate genes on the inactive X chromosome by regulating transcription factor accessibility to cis-acting DNA sequences. Furthermore, complete methylation of the HPRT promoter is not necessary to maintain the transcriptionally inactive state. The lack of a requirement for complete methylation emphasizes that either the overall density of CpG methylation or the position of methylated

sites in the 5' promoter region may be critical for maintenance of inactivation.

In Chapter 5, the methylation analysis of the human FMR1 gene trinucleotide repeat region was presented. Methylation analysis of the human FMR1 gene has demonstrated no methylation on the active X chromosome in normal and transmitting males, but in affected males and on the normal inactive X chromosome the trinucleotide repeat region is hypermethylated. Because a pattern of extensive DNA methylation of 5' CpG islands appears to be characteristic of the inactive X chromosome (36,86,110,119,120,126), the methylation analysis of the FMR1 gene suggests the hypermethylation of the trinucleotide repeat in fragile X males may be related to X chromosome inactivation. This is supported by the observation of hypermethylation at every CpG dinucleotide examined in the normal FMR1 gene on the inactive human X chromosome. Thus, transcriptional repression of the FMR1 gene in affected fragile X males may involve elements of X chromosome inactivation. Laird has postulated that hypermethylation and silencing of the FMR1 gene in affected fragile X males may be due to aberrant imprinting and failure of the inactive fragile X chromosome to reactivate during gametogenesis (54) in their mothers. However, this would require that the X chromosome carrying the fragile X mutation be selectively inactivated in the female germ line during embryogenesis since 100% of

premutations of the FMR1 gene above a threshold of 90 repeat units have been found to expand to the full mutation in oogenesis (26). Alternatively, the methylation patterns observed in the FMR1 gene may suggest that transcriptional repression of the FMR1 gene in fragile X males occurs by a mechanism similar to that used for transcriptional silencing of X-linked genes on the inactive X chromosome, but is not be due directly to the process of X inactivation.

Thus, in summary, this dissertation has investigated the mechanism(s) that regulate transcription of X-linked genes on the active and inactive X chromosome by X chromosome inactivation. These studies support the conclusion that transcriptional regulation of genes by X chromosome inactivation is probably secondary to differences in chromatin structure. Thus, future experiments should concentrate on the relationship of chromatin structure and X chromosome inactivation.

Investigation into the time course of 5-azacytidine reactivation on the inactive human HPRT gene suggests that demethylation and changes in nuclease sensitivity precede the initiation of transcription (99). In preliminary experiments, the human HPRT gene was studied by _in vivo_ footprinting during the 5-azacytidine reactivation process (Litt, Hornstra, Hansen, Gartler, and Yang; unpublished results). In initial experiments, the temporal appearance of the -91 footprint in the human HPRT gene coincides with

the reactivation of transcription (mRNA production). Thus, it appears that in order to reactivate the human HPRT on the inactive X chromosome with 5-azacytidine, demethylation and an increase in nuclease sensitivity occur before the expression of HPRT mRNA. These time course experiments examining the 5-azacytidine reactivation of the inactive HPRT gene emphasize the importance of chromatin structure in the process of X chromosome inactivation.

Future experiments to investigate the role of chromatin structure in X chromosome inactivation include the mapping of the human HPRT nuclease sensitive domain. Mapping of the nuclease sensitive domain may allow the identification of regulatory elements such as matrix attachment regions, boundary sequences, and locus control regions. Thus, the future research in X chromosome inactivation appears to depend on experiments which examine the regulation of chromatin structure.

REFERENCE LIST

1. Behe, M. and Felsenfeld, G. (1981). Effects of methylation on a synthetic polynucleotide: the B-Z transition in poly(dG-m5dC).poly(dG-m5dC). Proc. Natl. Acad. Sci. U.S.A. *78*: 1619-1623.

2. Bell, M.V., Hirst, M.C., Nakahori, Y., MacKinnon, R.N., Roche, A., Flint, T.J., Jacobs, P.A., Tommerup, N., Tranebjaerg, L., Froster Iskenius, U., and et al., (1991). Physical mapping across the fragile X: hypermethylation and clinical expression of the fragile X syndrome. Cell *64*: 861-866.

3. Ben-Hattar, J., Beard, P., and Jiricny, J. (1989). Cytosine methylation in CTF and Sp1 recognition sites of an HSV tk promoter: effects on transcription in vivo and on factor binding in vitro. Nucleic. Acids. Res. *17*: 10179-10190.

4. Bird, A. (1992). The essentials of DNA methylation. Cell *70*: 5-8.

5. Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*: 209-213.

6. Borsani, G., Tonlorenzi, R., Simmler, M.C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D., Lawrence, C., and et al., (1991). Characterization of a murine gene expressed from the inactive X chromosome. Nature *351*: 325-329.

7. Briggs, M.R., Kadonaga, J.T., Bell, S.P., and Tjian, R. (1986). Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. Science *234*: 47-52.

8. Brockdorff, N., Ashworth, A., Kay, G.F., Cooper, P., Smith, S., McCabe, V.M., Norris, D.P., Penny, G.D., Patel, D., and Rastan, S. (1991). Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. Nature *351*: 329-331.

9. Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P.,

Thirion, J.P., Hudson, T., and et al., (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member [published erratum appears in Cell 1992 Apr 17;69(2):385]. Cell *68*: 799-808.

10. Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafreniere, R.G., Xing, Y., Lawrence, J., and Willard, H.F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell *71*: 527-542.

11. Brown, C.J., Lafreniere, R.G., Powers, V.E., Sebastio, G., Ballabio, A., Pettigrew, A.L., Ledbetter, D.H., Levy, E., Craig, I.W., and Willard, H.F. (1991). Localization of the X inactivation centre on the human X chromosome in Xq13. Nature *349*: 82-84.

12. Brown, C.J. and Willard, H.F. (1990). Localization of a gene that escapes inactivation to the X chromosome proximal short arm: implications for X inactivation. Am. J. Hum. Genet. *46*: 273-279.

13. Buschhausen, G., Wittig, B., Graessmann, M., and Graessmann, A. (1987). Chromatin structure is required to block transcription of the methylated herpes simplex virus thymidine kinase gene. Proc. Natl. Acad. Sci. U.S.A. *84*: 1177-1181.

14. Carthew, R.W., Chodosh, L.A., and Sharp, P.A. (1985). An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. Cell *43*: 439-448.

15. Church, G.M. and Gilbert, W. (1984). Genomic sequencing. Proc. Natl. Acad. Sci. U.S.A. *81*: 1991-1995.

16. Cullen, C.R., Hubberman, P., Kaslow, D.C., and Migeon, B.R. (1986). Comparision of factor IX methylation on human active and inactive X chromosomes: implications for X inactivation and transcription of tissue-specific genes. EMBO J. *5*: 2223-2229.

17. Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic. Acids. Res. *11*: 1475-1489.

18. Dracopoli, N.C., Rettig, W.J., Albino, A.P., Esposito, D., Archidiacono, N., Rocchi, M., Siniscalco, M., and Old, L.J. (1985). Genes controlling gp25/30 cell-surface

molecules map to chromosomes X and Y and escape X-inactivation. Am. J. Hum. Genet. *37*: 199-207.

19. Dynan, W.S. and Tjian, R. (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell *35*: 79-87.

20. Edwards, A., Voss, H., Rice, P., Civitello, A., Stegemann, J., Schwager, C., Zimmermann, J., Erfle, H., Caskey, C.T., and Ansorge, W. (1990). Automated DNA sequencing of the human HPRT locus. Genomics *6*: 593-608.

21. Elgin, S.C. (1988). The formation and function of DNase I hypersensitive sites in the process of gene activation. J. Biol. Chem. *263*: 19259-19262.

22. Ellis, N., Keitges, E., Gartler, S.M., and Rocchi, M. (1987). High-frequency reactivation of X-linked genes in Chinese hamster X human hybrid cells. Somat. Cell Mol. Genet. *13*: 191-204.

23. Ellison, J., Passage, M., Yu, L.C., Yen, P., Mohandas, T.K., and Shapiro, L. (1992). Directed isolation of human genes that escape X inactivation. Somat. Cell Mol. Genet. *18*: 259-268.

24. Faisst, S. and Meyer, S. (1992). Compilation of vertebrate-encoded transcription factors. Nucleic. Acids. Res. *20*: 3-26.

25. Fried, M. and Crothers, D.M. (1981). Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic. Acids. Res. *9*: 6505-6525.

26. Fu, Y.H., Kuhl, D.P., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J., Holden, J.J., Fenwick, R.G.J., Warren, S.T., and et al., (1991). Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. Cell *67*: 1047-1058.

27. Fuscoe, J.C., Fenwick, R.G.J., Ledbetter, D.H., and Caskey, C.T. (1983). Deletion and amplification of the HGPRT locus in Chinese hamster cells. Mol. Cell Biol. *3*: 1086-1096.

28. Garner, M.M. and Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. Nucleic. Acids. Res. *9*: 3047-3060.

29. Garrity, P.A. and Wold, B.J. (1992). Effects of different DNA polymerases in ligation-mediated PCR: enhanced genomic sequencing and in vivo footprinting. Proc. Natl. Acad. Sci. U.S.A. *89*: 1021-1025.

30. Gartler, S.M. and Burt, B. (1964). Replication patterns of bovine sex chromosomes in cell culture. Cytogenetics *3*: 135-142.

31. Gartler, S.M. and Riggs, A.D. (1983). Mammalian X-chromosome inactivation. Annu. Rev. Genet. *17*: 155-190.

32. Gartler, S.M., Rivest, M., and Cole, R.E. (1980). Cytological evidence for an inactive X chromosome in murine oogonia. Cytogenet. Cell Genet. *28*: 203-207.

33. Grant, S.G. and Chapman, V.M. (1988). Mechanisms of X-chromosome regulation. Annu. Rev. Genet. *22*: 199-233.

34. Gross, D.S. and Garrard, W.T. (1988). Nuclease hypersensitive sites in chromatin. Annu. Rev. Biochem. *57*: 159-197.

35. Grumbach, M.M., Morishima, A., and Taylor, J.H. (1963). Human sex chromosome abnormalities in relation to DNA replication and heterochromatinization. Proc. Natl. Acad. Sci. U.S.A. *49*: 581-589.

36. Hansen, R.S., Ellis, N.A., and Gartler, S.M. (1988). Demethylation of specific sites in the 5' region of the inactive X-linked human phosphoglycerate kinase gene correlates with the appearance of nuclease sensitivity and gene expression. Mol. Cell Biol. *8*: 4692-4699.

37. Hansen, R.S. and Gartler, S.M. (1990). 5-Azacytidine-induced reactivation of the human X chromosome-linked PGK1 gene is associated with a large region of cytosine demethylation in the 5' CpG island. Proc. Natl. Acad. Sci. U.S.A. *87*: 4174-4178.

38. Hansen, R.S., Gartler, S.M., Scott, C.R., Chen, S., and Laird, C.D. (1992). Methylation analysis of CGG sites in the CpG island of the human FMR1 gene. Hum Mol. Genet. *1*: 571-578.

39. Harrington, M.A., Jones, P.A., Imagawa, M., and Karin, M. (1988). Cytosine methylation does not affect binding of transcription factor Sp1. Proc. Natl. Acad. Sci. U.S.A. *85*: 2066-2070.

40. Holler, M., Westin, G., Jiricny, J., and Schaffner, W. (1988). Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. Genes Dev. *2*: 1127-1135.

41. Hornstra, I.K. and Yang, T.P. (1992). Multiple in vivo footprints are specific to the active allele of the X-linked human hypoxanthine phosphoribosyltransferase gene 5' region: implications for X chromosome inactivation. Mol. Cell Biol. *12*: 5345-5354.

42. Huang, L.H., Wang, R., Gama Sosa, M.A., Shenoy, S., and Ehrlich, M. (1984). A protein from human placental nuclei binds preferentially to 5-methylcytosine-rich DNA. Nature *308*: 293-295.

43. Jalinot, P., Devaux, B., and Kedinger, C. (1987). The abundance and in vitro DNA binding of three cellular proteins interacting with the adenovirus EIIa early promoter are not modified by the EIa gene products. Mol. Cell Biol. *7*: 3806-3817.

44. Johnson, P. and Friedmann, T. (1990). Limited bidirectional activity of two housekeeping gene promoters: human HPRT and PGK. Gene *88*: 207-213.

45. Kaslow, D.C. and Migeon, B.R. (1987). DNA methylation stabilizes X chromosome inactivation in eutherians but not in marsupials: evidence for multistep maintenance of mammalian X dosage compensation. Proc. Natl. Acad. Sci. U.S.A. *84*: 6210-6214.

46. Kay, G.F., Penny, G.D., Patel, D., Ashworth, A., Brockdorff, N., and Rastan, S. (1993). Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation. Cell *72*: 171-182.

47. Keith, D.H., Singer Sam, J., and Riggs, A.D. (1986). Active X chromosome DNA is unmethylated at eight CCGG sites clustered in a guanine-plus-cytosine-rich island at the 5' end of the gene for phosphoglycerate kinase. Mol. Cell Biol. *6*: 4122-4125.

48. Kerem, B.S., Goitein, R., Richler, C., Marcus, M., and Cedar, H. (1983). In situ nick-translation distinguishes between active and inactive X chromosomes. Nature *304*: 88-90.

49. Keshet, I., Lieman Hurwitz, J., and Cedar, H. (1986). DNA methylation affects the formation of active chromatin. Cell *44*: 535-543.

50. Kim, S.H., Moores, J.C., David, D., Respess, J.G., Jolly, D.J., and Friedmann, T. (1986). The organization of the human HPRT gene. Nucleic. Acids. Res. *14*: 3103-3118.

51. Kovesdi, I., Reichel, R., and Nevins, J.R. (1987). Role of an adenovirus E2 promoter binding factor in E1A-mediated coordinate gene control. Proc. Natl. Acad. Sci. U.S.A. *84*: 2180-2184.

52. Kremer, E.J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S.T., Schlessinger, D., Sutherland, G.R., and Richards, R.I. (1991). Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. Science *252*: 1711-1714.

53. La Spada, A.R., Wilson, E.M., Lubahn, D.B., Harding, A.E., and Fischbeck, K.H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature *352*: 77-79.

54. Laird, C.D. (1987). Proposed mechanism of inheritance and expression of the human fragile-X syndrome of mental retardation. Genetics *117*: 587-599.

55. Lee, J.S., Woodsworth, M.L., Latimer, L.J., and Morgan, A.R. (1984). Poly(pyrimidine) . poly(purine) synthetic DNAs containing 5-methylcytosine form stable triplexes at neutral pH. Nucleic. Acids. Res. *12*: 6603-6614.

56. Levine, A., Cantoni, G.L., and Razin, A. (1992). Methylation in the preinitiation domain suppresses gene transcription by an indirect mechanism. Proc. Natl. Acad. Sci. U.S.A. *89*: 10119-10123.

57. Lewis, J. and Bird, A. (1991). DNA methylation and chromatin structure. FEBS Lett. *285*: 155-159.

58. Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell *69*: 905-914.

59. Lin, D. and Chinault, A.C. (1988). Comparative study of DNase I sensitivity at the X-linked human HPRT locus. Somat. Cell Mol. Genet. *14*: 261-272.

60. Liskay, R.M. and Evans, R.J. (1980). Inactive X chromosome DNA does not function in DNA-mediated cell transformation for the hypoxanthine

phosphoribosyltransferase gene. Proc. Natl. Acad. Sci. U.S.A. *77*: 4895-4898.

61. Lock, L.F., Melton, D.W., Caskey, C.T., and Martin, G.R. (1986). Methylation of the mouse hprt gene differs on the active and inactive X chromosomes. Mol. Cell Biol. *6*: 914-924.

62. Lock, L.F., Takagi, N., and Martin, G.R. (1987). Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. Cell *48*: 39-46.

63. Locker, J. and Buzard, G. (1990). A dictionary of transcription control sequences. DNA Seq. *1*: 3-11.

64. Lyon, M.F. (1992). Some milestones in the history of X-chromosome inactivation. Annu. Rev. Genet. *26*: 16-28.

65. Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, J., O'Hoy, K., and et al., (1992). Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. Science *255*: 1253-1255.

66. Maher, L.J., Wold, B., and Dervan, P.B. (1989). Inhibition of DNA binding proteins by oligonucleotide-directed triple helix formation. Science *245*: 725-730.

67. Maxam, A.M. and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. *65*: 499-560.

68. McBurney, M.W. (1988). X chromosome inactivation: a hypothesis. Bioessays *9*: 85-88.

69. Means, A.L. and Farnham, P.J. (1990). Transcription initiation from the dihydrofolate reductase promoter is positioned by HIP1 binding at the initiation site. Mol. Cell Biol. *10*: 653-661.

70. Meehan, R., Antequera, F., Lewis, J., MacLeod, D., McKay, S., Kleiner, E., and Bird, A.P. (1990). A nuclear protein that binds preferentially to methylated DNA in vitro may play a role in the inaccessibility of methylated CpGs in mammalian nuclei. Philos. Trans. R. Soc. Lond. Biol. *326*: 199-205.

71. Meehan, R.R., Lewis, J.D., and Bird, A.P. (1992). Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. Nucleic. Acids. Res. *20*: 5085-5092.

72. Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., and Bird, A.P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell *58*: 499-507.

73. Migeon, B.R., Shapiro, L.J., Norum, R.A., Mohandas, T., Axelman, J., and Dabora, R.L. (1982). Differential expression of steroid sulphatase locus on active and inactive human X chromosome. Nature *299*: 838-840.

74. Mohandas, T., Sparkes, R.S., and Shapiro, L.J. (1981). Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science *211*: 393-396.

75. Monaco, A.P. and Kunkel, L.M. (1988). Cloning of the Duchenne/Becker muscular dystrophy locus. Adv. Hum. Genet. *17*: 61-98.

76. Mueller, P.R. and Wold, B. (1989). In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. Science *246*: 780-786.

77. Nussbaum, R.L., Airhart, S.D., and Ledbetter, D.H. (1983). Expression of the fragile (X) chromosome in an interspecific somatic cell hybrid. Hum. Genet. *64*: 148-150.

78. Nussbaum, R.L. and Ledbetter, D.H. (1986). Fragile X syndrome: a unique mutation in man. Annu. Rev. Genet. *20*: 109-145.

79. Oberle, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M.F., and Mandel, J.L. (1991). Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science *252*: 1097-1102.

80. Ohno, S., Kaplan, W.D., and Kinosita, R. (1959). Formation of sex chromatin by a single-X chromosome in liver cells of Rattu norvegicus. Exp. Cell Res. *18*: 415-418.

81. Patel, P.I., Framson, P.E., Caskey, C.T., and Chinault, A.C. (1986). Fine structure of the human hypoxanthine phosphoribosyltransferase gene. Mol. Cell Biol. *6*: 393-403.

82. Patel, P.I., Nussbaum, R.L., gramson, P.E., Ledbetter, D.H., Caskey, C.T., and Chinault, A.C. (1984). Organization of the HPRT gene and related sequences in the human genome. Somat. Cell Mol. Genet. *10*: 483-493.

83. Pfeifer, G.P. and Riggs, A.D. (1991). Chromatin differences between active and inactive X chromosomes revealed by genomic footprinting of permeabilized cells using DNase I and ligation-mediated PCR. Genes Dev. *5*: 1102-1113.

84. Pfeifer, G.P., Steigerwald, S.D., Hansen, R.S., Gartler, S.M., and Riggs, A.D. (1990). Polymerase chain reaction-aided genomic sequencing of an X chromosome-linked CpG island: methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability. Proc. Natl. Acad. Sci. U.S.A. *87*: 8252-8256.

85. Pfeifer, G.P., Steigerwald, S.D., Mueller, P.R., Wold, B., and Riggs, A.D. (1989). Genomic sequencing and methylation analysis by ligation mediated PCR. Science *246*: 810-813.

86. Pfeifer, G.P., Tanguay, R.L., Steigerwald, S.D., and Riggs, A.D. (1990). In vivo footprint and methylation analysis by PCR-aided genomic sequencing: comparison of active and inactive X chromosomal DNA at the CpG island and promoter of human PGK-1. Genes Dev. *4*: 1277-1287.

87. Pieretti, M., Zhang, F.P., Fu, Y.H., Warren, S.T., Oostra, B.A., Caskey, C.T., and Nelson, D.L. (1991). Absence of expression of the FMR-1 gene in fragile X syndrome. Cell *66*: 817-822.

88. Povsic, T.J. and Dervan, P.B. (1989). Triple helix formation by oligonucleotides on DNA extended to the physiological pH range. J. Am. Chem. Soc. *111*: 3059-3060.

89. Razin, A. and Cedar, H. (1991). DNA methylation and gene expression. Microbiol. Rev. *55*: 451-458.

90. Riggs, A.D. (1990). DNA methylation and late replication probably aid cell memory, and type I DNA reeling could aid chromosome folding and enhancer function. Philos. Trans. R. Soc. Lond. Biol. *326*: 285-297.

91. Riley, D.E., Canfield, T.K., and Gartler, S.M. (1984). Chromatin structure of active and inactive human X chromosomes. Nucleic. Acids. Res. *12*: 1829-1845.

92. Riley, D.E., Goldman, M.A., and Gartler, S.M. (1986). Chromatin structure of active and inactive human X-linked phosphoglycerate kinase gene. Somat. Cell Mol. Genet. *12*: 73-80.

93. Rincon Limas, D.E., Krueger, D.A., and Patel, P.I. (1991). Functional characterization of the human

hypoxanthine phosphoribosyltransferase gene promoter: evidence for a negative regulatory element. Mol. Cell Biol. *11*: 4157-4164.

94. Riordan, J.R., Rommens, J.M., Kerem, B-S., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, L-C., Drumm, M.L., Iannuzzi, M.C., Collins, F.S., and Tsui, L-C. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science *245*: 1066-1080.

95. Rommens, J.M., Iannuzzi, M.C., Kerem, B-S., Drumm, M.L., Melmer, G., Dean, M., Rozmahel, R., Cole, J.L., Kennedy, D., Hidaka, N., Zsiga, M., Buchwald, M., Riordan, J.R., Tsui, L-C., and Collins, F.S. (1989). Identification of the cystic fibrosis gene: chromosome walking and jumping. Science *245*: 1059-1065.

96. Rousseau, F., Heitz, D., Biancalana, V., Blumenfeld, S., Kretz, C., Boue, J., Tommerup, N., Van Der Hagen, C., DeLozier Blanchet, C., Croquette, M.F., and et al., (1991). Direct diagnosis by DNA analysis of the fragile X syndrome of mental retardation. N. Engl. J. Med. *325*: 1673-1681.

97. Roy, A.L., Meisterernst, M., Pognonec, P., and Roeder, R.G. (1991). Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. Nature *354*: 245-248.

98. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

99. Sasaki, T., Hansen, R.S., and Gartler, S.M. (1992). Hemimethylation and hypersensitivity are early events in transcriptional reactivation of human inactive X-linked genes in a hamster x human somatic cell hybrid. Mol. Cell Biol. *12*: 3819-3826.

100. Seto, E., Shi, Y., and Shenk, T. (1991). YY1 is an initiator sequence-binding protein that directs and activates transcription in vitro. Nature *354*: 241-245.

101. Sherman, S.L., Jacobs, P.A., Morton, N.E., Froster Iskenius, U., Howard Peebles, P.N., Nielsen, K.B., Partington, M.W., Sutherland, G.R., Turner, G., and Watson, M. (1985). Further segregation analysis of the fragile X syndrome with special reference to transmitting males. Hum. Genet. *69*: 289-299.

102. Singer Sam, J., Grant, M., LeBon, J.M., Okuyama, K., Chapman, V., Monk, M., and Riggs, A.D. (1990). Use of a

HpaII-polymerase chain reaction assay to study DNA methylation in the Pgk-1 CpG island of mouse embryos at the time of X-chromosome inactivation. Mol. Cell Biol. *10*: 4987-4989.

103. Smale, S.T. and Baltimore, D. (1989). The "initiator" as a transcription control element. Cell *57*: 103-113.

104. Stout, J.T. and Caskey, C.T. (1985). HPRT: gene structure, expression, and mutation. Annu. Rev. Genet. *19*: 127-148.

105. Sutcliffe, J.S., Nelson, D.L., Zhang, F., Pieretti, M., Caskey, C.T., Saxe, D., and Warren, S.T. (1992). DNA methylation represses FMR-1 transcription in fragile X syndrome. Hum Mol. Genet. *1*: 397-400.

106. Sutherland, G.R., Gedeon, A., Kornman, L., Donnelly, A., Byard, R.W., Mulley, J.C., Kremer, E., Lynch, M., Pritchard, M., Yu, S., and et al., (1991). Prenatal diagnosis of fragile X syndrome by direct detection of the unstable DNA sequence. N. Engl. J. Med. *325*: 1720-1722.

107. Taylor, J.H. (1960). Asynchronous duplication of chromosomes in cultured cells of Chinese hamster. J. Biophys. Biochem. Cytol. *7*: 455-463.

108. The Huntingtons Disease Collaborative Research Group., (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntingtons disease chromosomes. Cell *72*: 971-983.

109. Therkelsen, A.J. and Brunnpeterson, G. (1967). Variation in glucose-6-phosphate dehydrogenase in relation to the growth phase and frequency of sex chromatin positive cell in cultures of fibroblasts from normal human females and a 48-XXXY male. Exp. Cell Res. *48*: 681-684.

110. Toniolo, D., Martini, G., Migeon, B.R., and Dono, R. (1988). Expression of the G6PD locus on the human X chromosome is associated with demethylation of three CpG islands within 100 kb of DNA. EMBO J. *7*: 401-406.

111. Venolia, L., Cooper, D.W., O'Brien, D.A., Millette, C.F., and Gartler, S.M. (1984). Transformation of the Hprt gene with DNA from spermatogenic cells. Implications for the evolution of X chromosome inactivation. Chromosoma *90*: 185-189.

112. Venolia, L. and Gartler, S.M. (1983). Comparison of transformation efficiency of human active and inactive X-chromosomal DNA. Nature *302*: 82-83.

113. Venolia, L., Gartler, S.M., Wassman, E.R., Yen, P., Mohandas, T., and Shapiro, L.J. (1982). Transformation with DNA from 5-azacytidine-reactivated X chromosomes. Proc. Natl. Acad. Sci. U.S.A. *79*: 2352-2354.

114. Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F.P., and et al., (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell *65*: 905-914.

115. Vincent, A., Heitz, D., Petit, C., Kretz, C., Oberle, I., and Mandel, J.L. (1991). Abnormal pattern detected in fragile-X patients by pulsed-field gel electrophoresis. Nature *349*: 624-626.

116. Wang, R.Y., Zhang, X.Y., Khan, R., Zhou, Y.W., Huang, L.H., and Ehrlich, M. (1986). Methylated DNA-binding protein from human placenta recognizes specific methylated sites on several prokaryotic DNAs. Nucleic. Acids. Res. *14*: 9843-9860.

117. Watt, F. and Molloy, P.L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev. *2*: 1136-1143.

118. Williams, T. and Tjian, R. (1991). Analysis of the DNA-binding and activation properties of the human transcription factor AP-2. Genes Dev. *5*: 670-682.

119. Wolf, S.F., Dintzis, S., Toniolo, D., Persico, G., Lunnen, K.D., Axelman, J., and Migeon, B.R. (1984). Complete concordance between glucose-6-phosphate dehydrogenase activity and hypomethylation of 3' CpG clusters: implications for X chromosome dosage compensation. Nucleic. Acids. Res. *12*: 9333-9348.

120. Wolf, S.F., Jolly, D.J., Lunnen, K.D., Friedmann, T., and Migeon, B.R. (1984). Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. Proc. Natl. Acad. Sci. U.S.A. *81*: 2806-2810.

121. Wolf, S.F. and Migeon, B.R. (1982). Studies of X chromosome DNA methylation in normal human cells. Nature *295*: 667-671.

122. Wolf, S.F. and Migeon, B.R. (1985). Clusters of CpG dinucleotides implicated by nuclease hypersensitivity as

control elements of housekeeping genes. Nature *314*: 467-469.

123. Yang, T.P. and Caskey, C.T. (1987). Nuclease sensitivity of the mouse HPRT gene promoter region: differential sensitivity on the active and inactive X chromosomes. Mol. Cell Biol. *7*: 2994-2998.

124. Yang, T.P., Singer Sam, J., Flores, J.C., and Riggs, A.D. (1988). DNA binding factors for the CpG-rich island containing the promoter of the human X-linked PGK gene. Somat. Cell Mol. Genet. *14*: 461-472.

125. Yen, P.H., Mohandas, T., and Shapiro, L.J. (1986). Stability of DNA methylation of the human hypoxanthine phosphoribosyltransferase gene. Somat. Cell Mol. Genet. *12*: 153-161.

126. Yen, P.H., Patel, P., Chinault, A.C., Mohandas, T., and Shapiro, L.J. (1984). Differential methylation of hypoxanthine phosphoribosyltransferase genes on active and inactive human X chromosomes. Proc. Natl. Acad. Sci. U.S.A. *81*: 1759-1763.

BIOGRAPHICAL SKETCH

Ian K. Hornstra was born in Merriam, Kansas, on September 26, 1962. I was raised in the Kansas City and attended Grandview High School. My father, Robijn K. Hornstra, M.D., is a psychiatrist at the state mental health facility in Kansas City and my mother, Mary Elizabeth Ritchie Hornstra, is a social worker and homemaker. I have one brother, Robijn, and one sister, Beth. After the completion of high school in 1980, I attended a six year B.A./M.D. program at the University of Missouri-Kansas City. I graduated medical school in May 1986 and went to Barnes Hospital at Washington University Medical Center in St. Louis, Missouri, for a residency in laboratory medicine. During my first year of residency, I decided to leave laboratory medicine and complete a year of internal medicine at Truman Medical Center in Kansas City, Missouri. After my year of internal medicine, I began graduate school at the University of Florida in the Department of Biochemistry and Molecular Biology in August 1988. After the completion of my dissertation, I will start a dermatology residency at Barnes Hospital of Washington University in July 1993.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Thomas P. Yang, Chairman
Assistant Professor of
Biochemistry and
Molecular Biology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Brian D. Cain
Assistant Professor of
Biochemistry and
Molecular Biology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Michael S. Kilberg
Professor of Biochemistry
and Molecular Biology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Robert J. Ferl
Professor of Horticultural
Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Harry S. Nick
Associate Professor of
Biochemistry and
Molecular Biology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Thomas W. O'Brien
Professor of Biochemistry
and Molecular Biology

This dissertation was submitted to the Graduate Faculty of the College of Medicine and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August 1993

Dean, College of Medicine

Gene Hemp
Dean, Graduate School

UNIVERSITY OF FLORIDA
3 1262 08554 6278